KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

PROCEEDINGS

SERIES B

PHYSICAL SCIENCES

VOLUME LXIV - No. 4

NORTH-HOLLAND PUBLISHING COMPANY - AMSTERDAM - 1961

The complete Proceedings consist of three Series, viz.:

SERIES A: MATHEMATICAL SCIENCES

SERIES B: PHYSICAL SCIENCES

SERIES C: BIOLOGICAL AND MEDICAL SCIENCES

Articles for these Series cannot be accepted unless formally communicated for publication by one of the members of the Royal Neth. Academy of Sciences.

IMPROVED DETECTION OF PHOSPHATIDES ON PAPER CHROMATOGRAMS WITH A MIXED STAINING SOLUTION. EFFECT OF CONTAMINATING SUBSTANCES IN THE PAPER. I

BY

H. G. BUNGENBERG DE JONG

(Communicated at the meeting of March 25, 1961)

1. *Introduction*

In a previous communication dealing with the use of Acid Fuchsin and uranyl nitrate in staining spots of phosphatides of the ampho-ion type (lecithin, kephalin, their lysoproducts and sphingo-myelin) a staining bath containing at the same time Brilliant Green [1] has been mentioned. With this mixed staining solution (0.01 % Brilliant Green, 0.006 % Acid Fuchsin, 0.2 % uranyl nitrate, 0.01 N HCl) apart from the red spots green spots appear which belong to components acidic in nature (Compare also [2]).

An example of its application to chromatograms of a commercial soybean phosphatide preparation has been given in [3].

A number of inconveniences adhere to the method originally given:

A) The green spots develop but slowly. After $^3/_4$ hours they become just visible and only after one night staining their maximum visibility is reached.

B) Because the green background colouration is intensified after drying, the green spots are hardly visible on the dry chromatogram. Hence it is advisable to circumline the green spots with a soft pencil on the washed and still wet chromatogram.

The final background colouration of the dried chromatogram increases with staining time. A modified mixed staining bath is therefore desirable, which stains faster and possibly with greater intensity (Improvements in this direction are given in *section* 3).

The green background colouration is due to contaminations of an acidic nature present in the chromatographic paper [1]. Weakly stained green spots on the chromatogram may be completely masked by this background colouration. The same may occur with weakly stained red spots, whereas stronger stained red spots may assume a purple colour.

By certain preliminary treatments of the paper, to be discussed in *section* 4, the contaminations causing a green background colouration are for the greater part removed. Masked spots thereby may become visible.

These treatments were found during a systematic investigation con-

cerning the removal of three classes of contaminations in the paper (acidic substances, aldehydic substances, ninhydrin positive substances, see [1]) by different liquids, as reviewed in the *sections* 5–8.

In section 9 results are reported which suggest that the acidic contaminations in the paper play a part in the degree of resolution of spots in chromatography.

2. *Methods and starting material*

The chromatographic paper used was SCHLEICHER and SCHÜLL, no. 2043*b*, either as such, or after a short treatment (10 seconds) with 2 % $SiCl_4$ in CCl_4, followed half an hour later by washing with distilled water.

$SiCl_4$–treated paper can be considered to be a silica-impregnated paper of low capacity. It is easily prepared and has the advantage that even large sheets (17 × 27 cm) show the same chromatographic properties on the whole sheet. For $SiCl_4$-treated papers see in general the series of communications mentioned in [4]. The variant employed in the present communication has been described in Part IV section 2 of the above mentioned series.

Chromatography is performed with di-isobutylketone – acetic acid – H_2O = 50:25:5, using either the small or the large slit-feeding apparatus, described in a former publication [5].

A good quality of technical soybean phosphatide preparation (still containing oil) was used. On the starting points (situated on a horizontal line 2 cm above the immersion line) 5 mm³ of a 2% solution of the above crude preparation in either isoamylalcohol-benzene = 1:1, or in chloroform-methanol = 80:20 was applied. Both mixtures dissolve the preparation completely and give no differences in the chromatograms.

For some experiments we brought up 5 mm³ of a reference mixture, a mixture containing lecithin, cephalin, its lysoproducts and sphingomyelin in suitable concentrations to give distinct, about equal sized, spots. This mixture has been used throughout the investigations in [4]. On the chromatograms these components are indicated by L, C, LL, LC and S. Another reference mixture containing only L, C, LL and S has also been used.

3. *Improvement of staining results by changes in the original composition of the mixed staining bath*

a) General

In the survey below we give, in row A, the composition of the original mixed staining bath [1]. On rows B and C follow the improved formulae, and on rows D and E the formulae of the washing liquids used in succession (first D, then E) as well after staining with B as with C.

Formula	Brilliant Green %	Acid Fuchsin %	uranyl nitrate %	HCl N	Acetic acid N
A	0.01	0.006	0.2	0.01	—
B	0.01	0.02	0.04	0.01	—
C	0.01	0.02	0.04	—	0.1
D	—	—	0.04	0.01	—
E	—	—	0.04	—	0.1

The original formula (A) and the newly proposed formulae (B and C)

contain next to Brilliant Green, the components (Acid Fuchsin and uranyl nitrate in acid medium) whereby spots of ampho-ionic phosphatides (L, C, LL, LC and S) are stained in a red colour.

This type of staining ("tricomplex staining") is in general diminished by salts other than uranylnitrate

It was found in [2] that in the presence of 0.3 N $(NH_4)_2SO_4$ in the staining bath the intensity of the green spots is considerably increased, but that of the red spots very markedly diminished. The influence just mentioned of salts on tricomplex staining explains this phenomenon.

If we assume that the increase of intensity of the green spots is a general effect of added salts, the desired effect (increased intensity of the green spots without loss of the intensity of the red spots) might be reached by simply increasing the uranyl nitrate concentration in the original formula A of the mixed staining bath. This expectation was however not fulfilled, since the intensity of the green spots also decreased. With $UO_2(NO_3)_2$ concentrations lower than in formula A, on the other hand, a marked increase of the intensity of the green spots took place. Hence both improved mixed staining baths, have a five times smaller $UO_2(NO_3)_2$ concentration than the original (compare B and C with A).

One cannot decrease the $UO_2(NO_3)_2$ concentration much farther, as here already the tricomplex staining shows a distinct decrease in intensity. To partly compensate the loss of intensity of the red spots the Acid Fuchsin concentration in the new formulae B and C is taken higher than in the original formula. (See survey).

The formulae B and C differ in that in B we retain 0.01 N HCl, whereas in C it is substituted by 0.1 N acetic acid. With C the green spots are better stained, but at the same time the green background colouration is also increased. With C a real gain in visibility of the green spots on the dried chromatograms is obtained only when after staining we wash first with the HCl containing washing liquid D to diminish the background colouration and then with E.

b) Storing the stock solutions. Influence of the pH and of temperature on the internal equilibrium of the staining solutions and on the staining results.

The freshly prepared mixed staining solutions are very dark and can not at once be used for staining. They brighten up at room temperature slowly (after a period in the order of 6 hours) as a result of the slow adjusting of the deep green colour of the Brilliant Green to a more or less yellowish green. The degree of the brightening up of the Brilliant Green is a (reversible) function of the pH.

In the case of the 0.01 N HCl (pH 2) containing mixed staining solution (B) the colour brightens to a red which is not much darker than that of an Acid Fuchsin solution of the same concentration as present in B. In the case of staining solution C, which has a higher pH (about 2.85)

the brightening up after reaching the internal equilibrium does not go as far. The solution remains markedly darker than a corresponding Acid Fuchsin solution.

We keep our staining solutions in stock in a refrigerator at about 5° C, to prevent mould growth. At this temperature brightening up goes farther than at room temperature (about 20° C). This, *of course*, is only striking in the case of the acetic acid containing staining solution. At room temperature, after reaching internal equilibrium, it has a purple tint, at 5° C it brightens up to a red somewhat darker than that of a corresponding Acid Fuchsin solution. This influence of the temperature is also reversible. The rate with which this change in the internal equilibrium takes place, once more is slow.

The knowledge of the above variables (pH and temperature) regulating the internal equilibrium, and of the fact that the equilibrium states are reached but slowly, has a direct bearing on the practice of staining. The darker the staining solution the more rapid and intense the green spots are stained, but on the other hand the background colouration is a darker green.

To obtain reproducible staining results we must start from staining solutions which have reached equilibrium, in our case from the solutions kept in stock in the refrigerator at least overnight.

Staining at this temperature (5° C) leads to poor results. Only some green spots appear and the colour intensity is slight, in particular with solution B. This slight staining obviously results from a too low concentration of the green form of Brilliant Green in the staining solution at 5° C. In practice, therefore, we stain at room temperature. We pour a portion of the staining bath in a glass dish and immediately immerse the chromatogram. The temperature of the staining solution (5° C) soon assumes room temperature, but as the internal equilibrium (see above) adjusts itself only slowly to the new temperature, the staining power increases while staining. As a rule we do not stain longer than 1 hour. When the chromatogram is left the night over in staining bath C, the background colouration has become so strong that the staining results are bad. Bath C has then assumed a violet colour and thus contains much more of the green form of Brilliant Green.

Thus we always stained with a fresh portion taken from the refrigerator which is discarded after use.

c) Staining in practice

The chromatogram is taken out of the chromatographic apparatus, is dried with an electrical hairdryer (Föhn), and is then left hanging in the air about $\frac{1}{2} - 1$ hour [1]).

[1]) When one postpones the staining to the following day, the staining proceeds slower. Besides it is not recommend to expose an already developed chromatogram a longer time to the air. Because of oxidation the intensity of staining of the red spots decreases.

It is then immersed in a fresh portion of staining bath taken from the refrigerator (5° C) for about 1 hour, blotted between filter paper, brought for 10 min. in washing liquid D, blotted, eventually washed another 10 min. in D [1]). After once more blotting, it is brought for 10 min. in washing liquid E, blotted and dried with a Föhn.

d) Staining results with chromatograms of crude soybean phosphatides on washed $SiCl_4$-treated paper.

From former work, using neutralized $SiCl_4$-treated paper [3] it appeared that there are at least 4 spots which may stain green, three of which are situated above the red cephalin spot and one below the red lecithin spot. Indicating them with α, β, γ and δ in downward direction, α is situated at the front. The acidic substance which is stained green is probably a fatty acid. Then follows β containing phosphorus but no choline, ethanolamine or inositol [2]).

According to newer observations in this laboratory the same applies for spot γ [3]). Possibly β and γ are "phosphatidic acids".

Spot δ is an inositol-phosphatide [2]). Its position differs hardly from the red spot belonging to lysolecithin which is also present in the crude phosphatide preparation.

It has been found that the same spots in the same sequence are obtained with the washed $SiCl_4$-treated paper used in the present communication. With the staining solution B, having the lower pH, the spots β, γ and δ are obtained as weakly green coloured spots on the dried chromatogram.

With the staining solution C, having a somewhat higher pH, spots β, γ and δ are stained green more strongly and a weakly green stained spot ε may be just visible below the other spots. Compare fig. 1. Spot ε contains inositol but no choline or ethanolamine [3]).

Spot α may be seen faintly, but on the dry chromatogram it generally disappears after some time, leaving a more or less white spot against the green background colouration, which, just below the front, is rather strong.

The small red LL spot is mostly masked by the green δ spot. Sometimes the LL spot may be discerned lying within the δ spot, as was the case in the chromatogram given in fig. 1.

e) Resorcinol or phenol containing staining solutions

Small additions of resorcinol or phenol to staining bath C markedly

---

[1]) Washing with D serve to remove as much as possible of the Brilliant Green to prevent the background from turning badly green after drying. This washing cannot be done much longer since the green spots also diminish thereby in intensity.

[2]) Personal communication of Dr. G. J. M. Hooghwinkel and Mr. G. M. van der Ent.

[3]) Personal communication of Mr. G. M. van der Ent.

increase the intensity of green staining of the spots α, β, γ, δ and ε so that they all become quite visible, despite the strengthened green background colouration. The concentrations should be kept low (1 % resorcinol

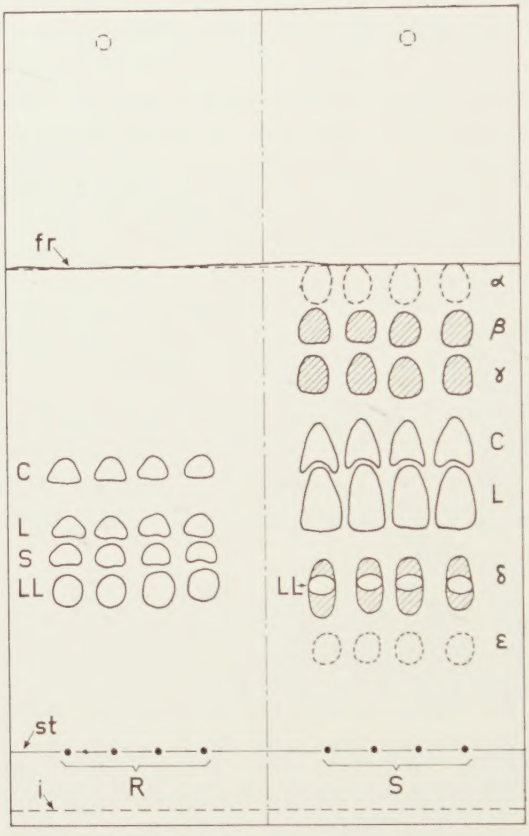

Fig. 1. Chromatogram on washed $SiCl_4$-treated paper of $\frac{1}{2}$ % reference mixture (5 mm³) on the left half of the sheet and of 2 % soybean phosphatides (5 mm³) on the right half. The left half has been stained with Acid Fuchsin + $UO_2(NO_3)_2$ in 0.01 N HCl, the right half with the improved mixed staining bath C. Red spots are circumlined only. Distinctly green spots are circumlined and shaded. The contours of very weakly stained (ε) or bleached (α) green spots are given as broken lines. The red spots on the left half of the chromatogram serve to identify the red spots originating from the soybean phosphatide preparation on the right half. For large red spots the position of the spot is characterized by its base; from this we conclude that cephalin (C), lecithin (L) and much less lysolecithin (LL) are present. For the spots denoted by α, β, γ, δ and ε see text.

or $\frac{1}{2}$ % phenol) lest the background colouration becomes too strong. Another danger of higher concentrations is that a flocculation occurs in the staining solution and darkly coloured floccules attach themselves to the chromatogram while staining.

This flocculation also occurs with 1 % resorcinol or $\frac{1}{2}$ % phenol but not during the first few hours after addition.

The procedure to be followed is as follows: after filling the glass tray

with a portion of the cold staining solution C (in stock in the refrigerator), one adds 1 gr. resorcinol of ½ gr. phenol to each 100 ml staining bath, dissolves this by rocking the tray and immerses the dried chromatogram. After a suitable staining time (1 hour or somewhat shorter) one twice washes with the washing solution D and once with washing solution E, after having added to each 100 ml of these solutions 1 gr. resorcinol or ½ gr. phenol.

The dried chromatogram keeps well when phenol has been used, but in the case of resorcinol the background turns into light brown and occasionally tends to develop brown flecks. It may be added that the washing liquids must contain resorcinol or phenol to keep the green spots well visible. When these are left out, the green spots diminish considerably in intensity while washing.

We have observed that a flocculated staining bath may still give a good colouration of the green spots after filtering off the flocculation. But it seemed that some time later a new flocculation occurs in the filtered solution. The possibility that a staining solution stable for a long time can be obtained has not been investigated. In practice we therefore use the resorcinol or phenol-containing staining bath only once.

A disadvantage of the use of phenol or resorcinol in the staining bath is the strong increase of the green background colouration.

A distinct advantage of these additions is only present when paper with reduced acidic contaminations is used. See section 4, paragraph c.

4. *Improvement of staining results by preliminary treatment of the paper to diminish the background colouration*

a) Distribution of the green background colouration on the stained chromatogram

The green background colouration on the chromatogram is not present with equal intensity between front and immersion line. A distinct zone is present at the front, while a much broader second zone of lesser intensity is present about half-way the distance front to immersion line. The limits of this latter green zone cannot be given accurately because of their vagueness. Calling the distance from the immersion line to the front 1.00, this zone has a width of about 0.2–0.25. Weak green spots lying in this zone may become hardly visible, and weak red spots also, whereas more strongly stained red spots may assume a violet colour.

Better staining results can therefore be expected if, by certain preliminary treatments, the contaminations in the paper which stain green, could be removed. In section 5 a short survey will be given of attemps in this direction. Although we have not succeeded in removing all of the green background, it proved possible to diminish markedly the intensity of the green zone halfway front to immersion line, which is particularly troublesome. Some treatments are given in the next paragraph.

b) Treatments of the paper leading to a diminished green background colouration

Depending on the procedure to be followed, we may divide the preliminary treatments into three classes:

1) The paper sheet (17×28 cm) is bathed in a suitable liquid, which extracts contaminations.

2) The paper is placed in the large slit-feeding apparatus and a suitable liquid is allowed to rise into the paper overnight.

3) The same as sub 2), followed after drying by a second treatment with another liquid in the slit-feeding apparatus, also overnight.

Procedure 1) is the simplest, only it consumes relatively more liquid. A sheet of 17×27 cm washed $SiCl_4$-treated paper is simply immersed in 100 ml liquid in a glass tray and covered with a glass plate. After one or more days the sheet is removed and dried in air. A reasonable decrease of the green background colouration is already obtained when the extraction is performed overnight with ethanol 95 %.

Procedure 2) carried out using our slit-feeding apparatus, consumed only some 10 ml liquid for a sheet of 17×28 cm. As stated, the liquid ascends overnight. The time necessary for the front to reach the upper rim of the paper 27.5 cm above the immersion line, is much shorter. But after this upper rim is reached, the ascending movement of the liquid through the paper in upward direction does not stop at once, but continues for some time. The liquid content in the paper decreases in the direction of the front. When the front has reached the upper rim, the paper continues to absorb liquid as a result of the tendency to reach an equal liquid content in the whole of the wetted paper. Thus extraction and transportation of acid contaminations in the paper in upward direction continues for some time after the front has reached the upper rim of the paper.

After drying in air, such sheets can be used for chromatography in the large slit-feeding apparatus. We prefer to allow the front to rise to a height much less than 27.5 cm, as the extracted and transported contaminations which stain green are present in the upper part of the paper. Thus the front is allowed to ascend only to 18 cm above the immersion line. Reasonable results have been obtained with methanol, ethanol, acetone and acetic acid.

Procedure 3)

Reasonably good results have been obtained by using as first liquid

ethanol and as following liquid cellosolve. Still better results are obtained by using 86 % n. butanol as first liquid, followed by cellosolve.

c) Staining results with chromatograms of crude soybean phosphatides

Well visible green λ and ε spots are obtained when chromatograms on washed $SiCl_4$ treated paper pretreated according to procedure 3 are stained with the mixed staining bath to which 1 % resorcinol has been added.

Fig. 2 gives an example. However, as a result of the added resorcinol, the front zone is stronger stained but the λ spots embedded in this zone can be discerned without difficulty.

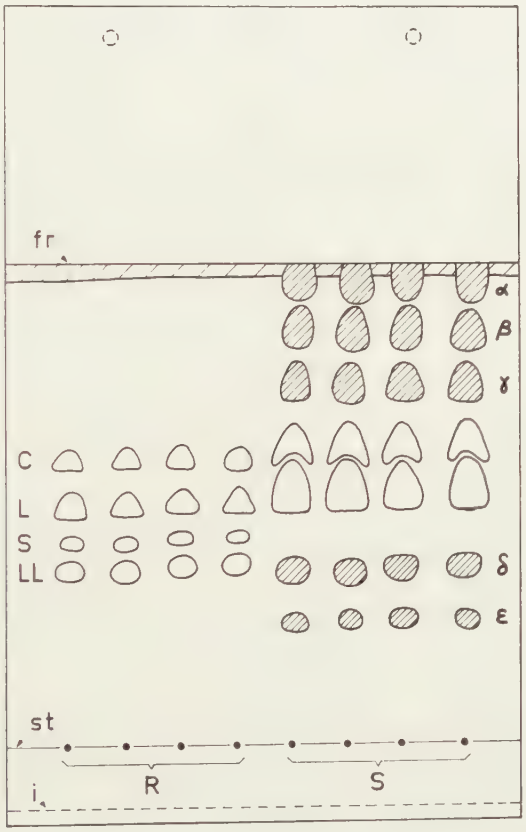

Fig. 2. Chromatogram as in fig. 1, however, on washed $SiCl_4$-treated paper, treated before chromatography according to procedure 3 (ethanol followed by cellosolve). The whole chromatogram has been stained with staining bath C, to which 1 % resorcinol had been added. Further description as in fig. 1.

Approximately the same result are obtained using paper pretreated with ethanol according to procedure 2, and staining with the resorcinol containing mixed staining bath.

5. *Comparison of some liquids as to their ability to remove contaminations present in the paper which stain with Brilliant Green, give the Schiff reaction, or the ninhydrin reaction*

a) Method

The presence of three classes of contaminants in chromatographic paper has already been discussed in a former communication [1]. Here we will also use the terms "Brilliant Green substances", "Aldehydic substances" and "Ninhydrin positive substances", and we will report experiments with untreated Schleicher and Schüll no 2043*b*. With each liquid three strips (4.5 × 22 cm) have been run in a small slit-feeding apparatus to a front height of 18 cm above the immersion line. After drying in the air one strip was stained with the mixed staining bath C. The second strip served for the Schiff reaction and the third for the ninhydrin reaction. We then compared the results.

Intensity of colouration	Staining with Brilliant Green		Schiff reaction	Ninhydrin reaction
	frontzone	background	frontzone	frontzone
strong	acetic acid	50:25:5	$CHCl_3 + CH_3OH$	H_2O
	$CHCl_3 + CH_3OH$ (4:1)		acetic acid	acetic acid
			$CHCl_3$	
moderately strong	50:25:5 [1]	benzene	benzene	
	acetone	$CHCl_3$	CCl_4	
	CCl_4	H_2O	50:25:5	
	benzene	CCl_4		
weaker	$CHCl_3$	di-isobutyl ketone	di-isobutyl ketone	CH_3OH
	di-isobutyl ketone	acetone	acetone	
very weak or not stained	H_2O [2]	methanol	H_2O	all other liquids mentioned
		$CHCl_3 + CH_3OH$ [2]		
		acetic acid [3]		

b) Results regarding Brilliant Green staining

In columns 2 and 3 of the above survey the liquids investigated have been classed in four groups according to two criteria, namely to the intensity of the green zone at the front and to the general aspect of the green background colouration between

[1] The mobile phase used throughout in chromatography with $SiCl_4$ — treated paper consisting in di-isobutylketone – acetic acid — H_2O = 50:25:5.

[2] The green front zone is absent here.

[3] On the lower end of the strip, see section 6.

front zone and immersion line. Two remarks must be added: 1) The intensities of the background colouration are much lower than the intensities which may be reached by the front zone. Thus the designations strong, moderately strong, weaker, very weak or not stained are relative only.

2) By classing the liquids as to the background colouration we give only the general impression, disregarding the fact that in some cases zones occur (for this see section 6).

c) Results of the Schiff-reaction

In column 4 of the survey we give only the result of the Schiff reaction of the zone at the front; there is not much difference in the intensity of the Schiff reaction of the background.

In the case of acetic acid a broad zone at the front gave an extremely intense red colour at the first moment in the Schiff reaction, but this disappeared by rinsing with SO_2-water.

We consider this initial colour reaction not as a reaction on aldehydic substances. The rinsed strip shows at the front a narrower zone with a still strong normal colour for the Schiff reaction. The background colouration in the case of methanol is discussed in section 6.

d) Results of the ninhydrin reaction

The results are given in column 5 of the survey for the zone at the front, the background colouration here being slight in general and not giving distinct differences between the liquids investigated.

e) Comparison of Brilliant Green staining at the front zone and of the background

When we compare columns 2 and 3 in the survey, we generally find that when a strong green front zone is obtained, the background colouration is less. The reverse also applies; a stronger background colouration is correlated with a more weakly coloured zone at the front.

f) Comparison of Brilliant Green staining and Schiff reaction of the zone at the front

When we compare columns 2 and 4 of the survey, we observe a parallelism. Liquids which give a strong green front zone (acetic acid, $CHCl_3$ + CH_3OH (4:1)) give also a strong Schiff reaction at the front. On the other hand with H_2O both reactions are negative at the front.

In a former communication [1] such parallelism was already observed regarding the successive terms of the homologous series of the normal primary alcohols.

g) Comparison of the ninydrin reaction at the front with the Brilliant Green staining and the Schiff reaction at the front

In the communication [1] just quoted, we found an anti-parallelism between the ninydrin reaction and the two other reactions as regards the successive terms of the homologous series of the normal primary alcohols. Comparing column 5 with columns 2 and 4 we find too an anti-parallelism. It is only H_2O which gives a distinct ninhydrin reaction at the front, a weak reaction is shown in the case of methanol, while all other liquids showed no ninhydrin reaction at the front.

6. *Indications for a plurality of acidic and aldehydic contaminations in the paper*

a) Zones of acidic contaminations below the front

In the aforementioned communication [1] we already obtained, by means of paper electrophoresis indications, that in the front zone there is not just one

acidic substance which stains green with Brilliant Green, but several. The plurality of "Brilliant Green substances" comes also to the fore in the uneven distribution of the green background colouration which was observed in the present investigation with a number of liquids.

Chloroform-methanol (4:1)

Adjacent to the strong green zone at the front, a weaker green coloured zone stretches downward to about halfway the distance front to immersion line, see Fig. 3, A.

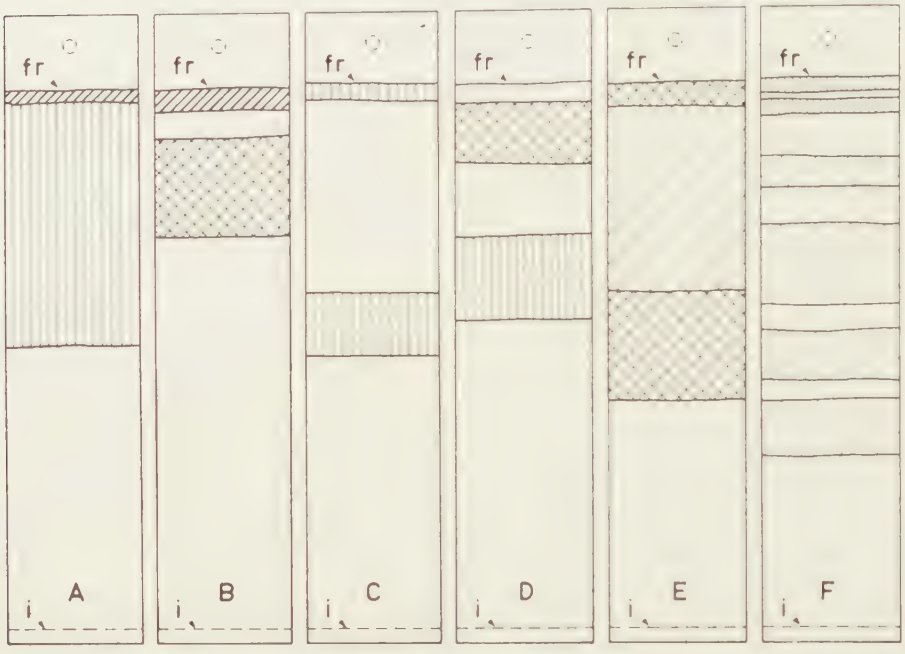

Fig. 3. Zones of Brilliant Green stainable contaminations in untreated S + S 2043*b* paper after ascension with various liquids.

Acetic acid

A short distance beneath the strong green zone at the front begins a weaker green stained zone, which ends at about 0.7 (the front taken as 1.0 and the immersion line as 0.0). See Fig. 3, B.

Di-isobutylketone

Apart from the not very strong green zone at the front, one finds a weakly coloured green zone about half-way the distance front to immersion line. See Fig. 3, C.

Methanol

No green zone is present at the front, but somewhat below the front a distinct green zone is present. A second weaker green zone is found somewhat lower. See Fig. 3, D.

It is clear that the above examples point to at least two different acidic contaminations in the paper.

Di-isobutylketone + acetic acid + H_2O = 50:25:5

Here we find a green zone at the front, a broad, distinctly green zone about the middle of the distance front to immersion line, whereas between this zone

and the front zone the paper is stained more strongly than below the middle zone. This points to the presence of at least three acidic contaminations in the paper. See Fig. 3, E.

Acetone

Apart from the moderately coloured green zone at the front, one finds a number of very weakly coloured green zones. Possibly there are five such zones. See Fig. 3, F.

This would mean, that there are at least six acidic contaminations in the paper.

b) Zones of aldehydic contaminations below the front

The strip in which methanol had ascended showed three zones in which the weak violet background colouration was strong. A relatively strong zone occurred a short distance below the front. A more weakly coloured zone was present about half-way the distance front to immersion line. Between the two abovementioned zones a third zone was just visible. Compare B in Fig. 4.

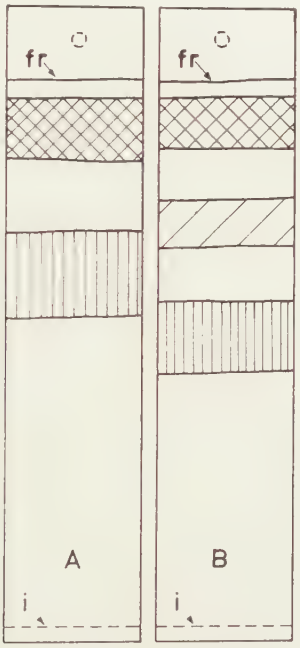

Fig. 4. Strips in which methanol has ascended. Left strip stained with the mixed staining bath. Right strip, appearance after the Schiff reaction.

It is interesting to compare this strip with the methanol strip which had been stained with the mixed staining bath. We therefore reproduce in Fig. 4, A the strip already given in Fig. 3, D.

Comparison then shows that zones of acidic and aldehydic contaminations partly overlap and partly alternate.

(To be continued).

IMPROVED DETECTION OF PHOSPHATIDES ON PAPER CHROMATOGRAMS WITH A MIXED STAINING SOLUTION. EFFECT OF CONTAMINATING SUBSTANCES IN THE PAPER. II

BY

H. G. BUNGENBERG DE JONG

(Communicated at the meeting of March 25, 1961)

7. *Some further particulars on the plurality of acidic contaminations in untreated and $SiCl_4$-treated paper. Influence of heating*

a) Method

In this section we give some results on untreated and on washed $SiCl_4$-treated paper run in the large slit-feeding apparatus with di-isobutylketone - acetic acid H_2O = 50:25:5. The sheets (17 × 27 cm) were dried with a cold air stream after chromatography and then cut in 4 equally broad strips. Two strips of each sheet are used for the experiments, the two other strips of each sheet served for other purposes, not discussed here.

Of the two strips from the sheet of untreated paper, one was heated half an hour at about 90°, the other served as blank. The same was done with two strips of the $SiCl_4$-treated paper. We then stained them in the mixed staining bath. The results are given in Fig. 5.

b) Untreated paper

The unheated strip (fig. 5, A) shows, in accordance with the results given in section 6, a green zone at the front and a weakly stained green zone about half way the distance from front to immersion line.

The strip which was heated (fig. 5 B) shows a more intense zone at the front, and below it two distinctly visible green zones connected by a less green zone. Thus previous heating increases in general the intensity of the green staining of acidic contaminations. For the appearance of a second green zone two explanations are possible:

i) The corresponding zone of acidic substances are present also in the unheated paper; heating removes something which prevents staining.

ii) The acidic substances are not present in the unheated paper, but are formed from other substances present in this zone by heating. The appearance of the second green zone after heating once more shows the plurality of contaminations, moving with chromatography in the paper.

c) Washed $SiCl_4$ treated paper

When the strips of the $SiCl_4$-treated paper are compared to the strips of the untreated paper, one observes that in general the intensity of staining on $SiCl_4$-treated paper is greater. The influence of heating is partly the same as with untreated paper; partly much more complex. On the unheated strip (fig. 5, C) we observe, as on untreated paper, a green zone at the front and a broad, less intense green zone about half-way the distance from front to immersion line.

On the heated strip, as for untreated paper, a second green zone appears (fig. 5, D).

A remarkable influence of heating manifests itself in the front zone. Whereas in the untreated strip it shows a simple evenly stained green zone, the heated strip

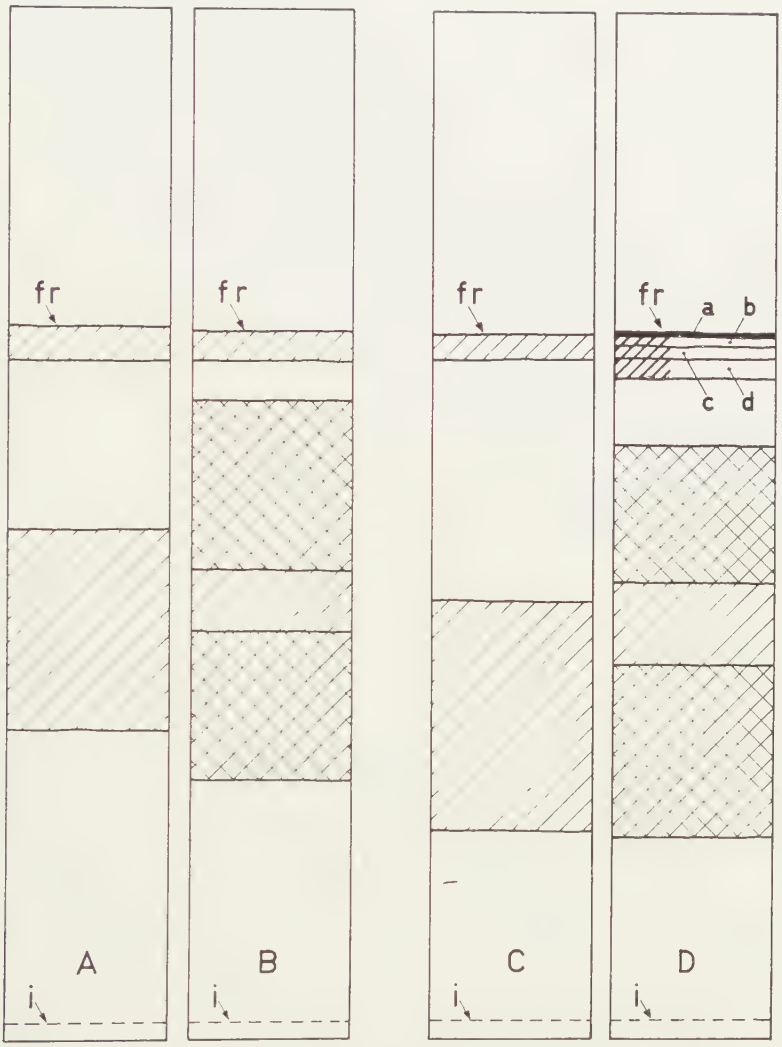

Fig. 5. Staining results after ascension of the mobile phase 50:25:5 in non-impregnated paper (A and B) and in washed $SiCl_4$-treated paper (C and D). Before staining, strips B and D have been heated $\frac{1}{2}$ hr at 90°, strips A and B have not been heated. The front zone in strip D consists of four adjacent zones.

has at the front four distinct adjacent zones, denoted in fig. 5 by a, b, c and d.

At the frontline we find a narrow blue zone (a), then follows in downward direction a light green zone (b), a purple zone (red before drying) (c) and a stronger green zone (d).

Heating thus reveals the multiplicity of contaminating substances present below the front, arranged in adjacent zones. The blue colour at the front line is probably a metachromatic staining of an acidic substance by Brilliant Green. The purple (when still wet: red) zone points to a substance which is either basic in character,

staining with the Acid Fuchsin in acid medium, or amphoteric (phosphatide?) giving the tricomplex staining with Acid Fuchsin + uranyl nitrate. [1])

d) Discussion

From the above experiments we may draw a practical conclusion. To remove the mobile phase from a just finished chromatogram hot air from a hairdryer should be applied for a short time only, lest the colouration of the background and of the front zone increases.

8. *An efficient method to remove from untreated paper Brilliant Green substances*

In section 5 it was found that the most strongly coloured green front zone was obtained with acetic acid.

Two other liquids give comparable strong green front zones after staining, namely 86 % (by volume) n. butanol and cellosolve.

It was tried to remove the Brilliant Green substance by successive ascension of the three liquids mentioned.

There are six sequences:

1) acetic acid — 86 % n. butanol — acetic acid
2) acetic acid — cellosolve — 86 % n. butanol
3) 86 % n. butanol — acetic acid — cellosolve
4) 86 % n. butanol — cellosolve — acetic acid
5) cellosolve — acetic acid — 86 % n. butanol
6) cellosolve — 86 % n. butanol — acetic acid

Strips (4.5 × 21 cm) of untreated S + S paper were run in the small slit-feeding apparatus over night each time with the first named liquid. After drying in air they were run the following day to a front height of 18 cm above the immersion line with the second liquid mentioned. After drying once more in air they were run to a front height of 16 cm with the third liquid mentioned.

After once again drying in air the six strips were stained in the mixed staining bath C (survey in section 3). After washing in washing-bath D and E, all six strips showed below the last front line (16 cm) no green background colouration, the whitest colour being present when the abovementioned sequence no. 3 had been used.

With the sequences no. 1 and 2, no. 3 and no. 5 we found at the upper margin, at the 18 cm front line, and at the 16 cm front line strong green front zones (in the sequences nos. 1 and 2, the whole zone between margin and the 18 cm line was strongly coloured with a bluish tint).

It follows that, every time two ascensions: acetic acid — 86 % n. butanol; acetic acid — cellosolve; 86 % butanol — acetic acid; cellosolve — acetic acid, do not remove all Brilliant Green substance. With the sequences no. 4 and no. 6 we found only two green zones, namely at the upper margin and at the 18 cm front line, but no green front zone at 16 cm.

It follows that both successions: 86 % n. butanol — cellosolve and cellosolve — 86 % n. butanol, effectively remove the Brilliant Green substances, so that with acetic acid no more Brilliant Green is left to give a green front zone.

On these results we based the pretreatment of $SiCl_4$-treated paper mentioned in section 4 as procedure 3. After washing out and drying the background in this case,

[1]) Though proteins are stained red with Acid Fuchsin in acid medium, it seems improbable that the red zone at the front indicates protein. As far as we know proteins in chromatography with our mobile phase never move with the front. In many cases they remain on the starting point.

however, is not quite white, probably as a result of the strong adsorptive power of SiO_2 in the impregnated paper.

9. *Indication that contaminations of acid nature moving during chromatography in the paper have influence on the position of the spots*

a) Method

We already saw in section 5 and 6, that our mobile phase (di-isobutylketone – acetic acid – $H_2O = 50:25:5$), when ascending in the paper, mobilizes and transports substances which can be stained green. We may mainly discern two groups of these, one ascending with the front and a second moving about half-way the distance from front to immersion line. Indications that the latter group influences the position of the spots on the chromatogram will now be discussed. In the following exposition of the method we will, therefore, pay no attention to the first mentioned group.

A large slit-feeding apparatus is used for the experiments, allowing to run paper sheets of 24×24 cm. If the mobile phase is allowed to ascend to 18 cm front height and afterwards the sheet is stained with the mixed staining bath, a staining pattern would be obtained as shown by I in fig. 6. We discern four zones a, b, c and d in the paper.

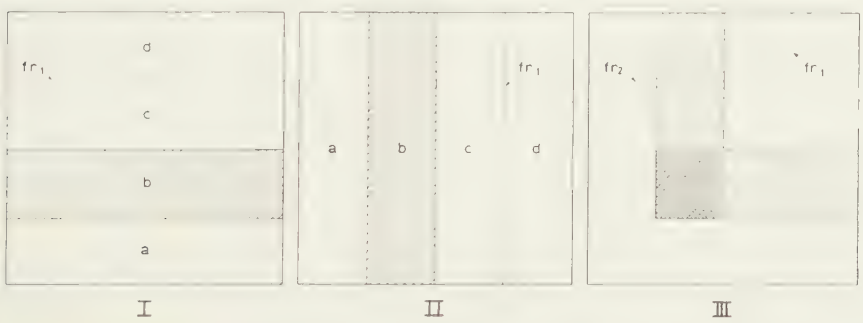

Fig. 6. Diagrams illustrating the method followed in section 9.

The considered group of acidic contaminations is for the greater part no longer present in zone a. The part extracted is present in zone b, whereas in zone c these acidic contaminations are still present unchanged. In zone d, which has only been exposed to the vapours of the mobile phase, all acidic contaminations are, of course, present unchanged.

In practice we do not stain the chromatogram after the first ascension with the mobile phase. We suspend the chromatogram in air for a week [1]) and then fasten it anew in the slit-feeding apparatus, now, however in a position turned clockwise over 90°. In this position the zones a, b, c and d now stand vertical. See II in fig. 6.

Before bringing the sheet in this turned position on its place, we

[1]) To evaporate from the paper a remnant of the mobile phase. The phenomenon to be described is, however, in principle the same after only one day suspension in air.

apply on the starting points on a horizontal line 5 mm^3 of the 2 % solution of the soybean phosphatide preparation. Thereafter we allow the mobile phase to ascend to a front height of 18 cm. After drying the chromatogram, it is stained with the mixed staining solution.

The green staining pattern in the paper background due to the Brilliant Green substances then will be as given in III in fig. 6.

An answer to the question whether Brilliant Green substances influence the chromatographic displacement of spots, can be seen at once from the position of the spots on the chromatogram developing from the starting points lying in the zones a, b, c and d respectively. Similar experiments have been performed with 5 mm^3 of the 0.5 % reference mixed applied on the starting point. In this case we stained only with Acid Fuchsin $+UO_2(NO_3)_2+0.01$ N HCl. Here of course the zones in which the Brilliant Green substances are present are not visible on the stained chromatogram, but their position is known from the chromatograms of the soybean phosphatide preparation.

b) Chromatograms on non-impregnated paper

Fig. 7 gives the chromatogram for the soybean phosphatide preparation obtained in the way as discussed in the preceding paragraph. The approximate limits of the green staining pattern in the paper background are given by broken lines. This pattern has indeed the shape as expected above. See diagram III in fig. 6.

Besides, we find a strongly green, comma shaped spot near to the crossing point of front I and front II (hatched in fig. 7) which is due to an accumulation of Brilliant Green substances originally present along front I. With the second ascension this has been transported upwards with the ascending front II [1]).

The spots on the chromatogram have been circumlined, they are stained green (α, β, γ, δ and ε) or red (the composite L+C spot and the small LL spot).

Spots or parts of spots which on the dried chromatogram appeared as white, against the more or less green-stained surroundings, are indicated by broken lines.

This occurs in the row of α spots lying at front II. Here we have already an indication that with the first ascension one or more contaminations have been transported, which counteract the staining green of the α spots or lead to a more or less rapid bleaching away of the green colour.

The α spots originating from the three starting points to the left are stained green permanently. These three starting points lie in zone a of

[1]) In the figs. 7, 8, 9 and 10, front II lays somewhat behind in the zone between front I and the right margin of the paper. The slightly lower rate of ascension is obviously due to the fact that during the first ascension this zone had not been wetted by the mobile phase.

diagram III of fig. 6, that is, in the zone which has been practically freed from that class of Brilliant Green substances which move about half-way the distance front to immersion line. The α spots which

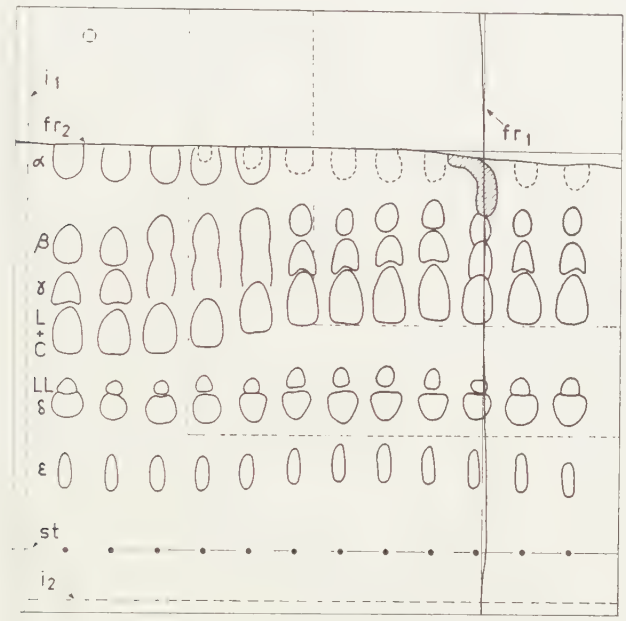

Fig. 7

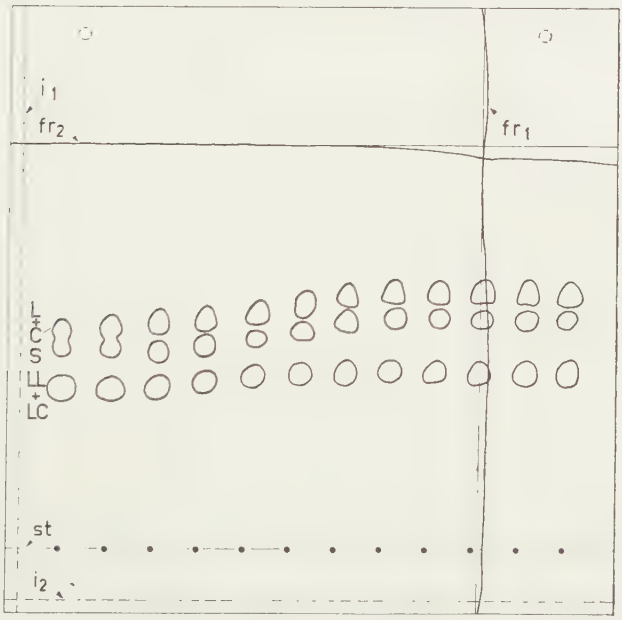

Fig. 8

Fig. 7. Chromatogram of soybean phosphatides on non-impregnated paper. Preceding chromatography to front II, the mobile phase was allowed to ascend to front I perpendicular to the direction of chromatography.

Fig. 8. Chromatogram of the reference mixture on non-impregnated paper. Further description as in fig. 7.

originated from the starting point 4, 5 and 6 (reckoned from the left), show the transition from green to white. This occurs in zone b of diagram III of fig. 6.

The α spots originating from the starting points more to the right appear as white spots only [1]). These starting points are situated in zones c and d of diagram III of fig. 6, in which the class of Brilliant Green substances moving about halfway the distance from front to immersion line are still present unchanged.

We now turn to the remaining rows of spots. We observe that the spots LL, δ and ε lie practically on horizontal lines. For these spots it does not matter whether the starting points lie in the zones a, b, c or d of diagram III of fig. 6.

For the three higher-situated rows of spots L+C, γ and β this does not apply.

In zone a, where the Brilliant Green substances are for the greater part no longer present at the start of the second ascension, they stand on lower levels than in zones c and d, in which this group of Brilliant Green substances is still present [2]). The transition from the lower situated to the higher-situated levels occurs in zone b, of diagram II of fig. 6. In this transition zone the contours of the spots β and γ become difficultly visible as a result of the green background coloration present here.

We now turn to an analogous chromatogram on which we had on the starting points applied 5 mm³ of the $\frac{1}{2}$ % Reference mixture, containing lecithin (L), cephalin (C), their lysoproducts (LL and LC) and sphingomyelin (S).

On non impregnated paper we obtain only three spots, the upper spot containing L and C, the middle spot S and the lower spot containing LL and LC. The chromatogram after staining with Acid Fuchsin $-UO_2(NO_3)_3-0.01$ N HCl is given in fig. 8. We observe here that the rows of C+L spots and of S spots behave in a similar way as in fig. 7 applied for the C+L, the γ and β spots. The levels on which they are situated to the left of the chromatogram lie distinctly lower than on the right half of the chromatogram, the transition to these higher levels occurring once more in zone b of diagram III in fig. 6.

[1]) Because of the accumulation of "front-Brilliant Green substances" at the intersection of front I and front II (see the comma shaped hatched spot), spot α here is not visible.

[2]) The train of spots L + C, γ and β, originating from the third starting point reckoned from the right, has not been developed fully, because of the "front Brilliant Green substances" (the hatched comma-shaped spot) present in quantity at the intersection of front I and front II. Signs of this hindrance are 1) the fact that L + C, γ and β form here a closed group and 2) the lower situated top of the L + C spot. The base of the latter spot lies, however, on the same horizontal level as those of the other L + C spots in the right half of the chromatogram.

The row of spots LL+LC does not show this transition to a higher level [1]).

Combining the results of the chromatograms fig. 7 and fig. 8, we find that on non-impregnated paper the position of the spots β, γ, L+C and S is influenced by certain contaminations in the paper. When these contaminations are still present the said spots lie higher than when they are absent.

The position of the lower situated spots LL, LC, δ and ε on the other hand is not influenced by the presence or absence of these contaminations.

c) Chromatograms on washed $SiCl_4$-treated paper

Here too we made one chromatogram with soybean phosphatides applied and one with the reference mixture applied. See fig. 9 and fig. 10. Differences of these chromatograms with the analogous chromatograms on non-impregnated paper (fig. 7 and fig. 8), due only to impregnation, are the following:

1) The L+C spot is now separated into two separate spots (fig. 9 and fig. 10).
2) The LL+LC spot is now divided to an separate LL spot, whereas the LC spot forms with S a composite spot (fig. 10).
3) The LL spot, which on non-impregnated paper lies a short distance above spot δ, now coincides with the δ spots and is masked by the green colour of the δ spot (fig. 9).

We now turn to the question whether contaminating substances in the paper which have been transported with the preceding ascension here have an analogous influence as on non-impregnated paper. First consider in fig. 9 the spots λ at front II. As in the case of non-impregnated paper (fig. 7), the α spots situated in the zones c and d in diagram III of fig. 6, appear as white spots. In zone b a transition takes place, the λ spot now being partly green and partly white. This state of the λ spots remains in zone a. This points to a less effective removal of the contamination counteracting the total staining of the α spot.

Now turning to the other rows of spots in fig. 9, we find in principle the same as for the analogous chromatogram on non-impregnated paper (fig. 7).

Once more the accumulation of "front Brilliant Green substances"

[1]) The spots on this row do not stand on a strictly horizontal line, but on a line slightly sloping upwards to the right. This may also occur sometimes with the spots on chromatograms which have been run in the ordinary way (thus without previous ascension in a direction at right angles). Spots not on horizontal lines are probably caused by one-sided radiation. Though chromatography is always performed in a dark room of constant air temperature, one-sided radiation may still be present, for instance to the wall or the door of the room. Recently it has been found that by shielding of the chromatographic apparatus by a cage of aluminum plate, the occurence of spots not on horizontal lines becomes much less frequent.

near the crossing point of front I and front II (hatched comma-shaped spot) compresses here the train of spots β, γ, C and L, originating from the third starting point reckoned from the right. Here we observe that

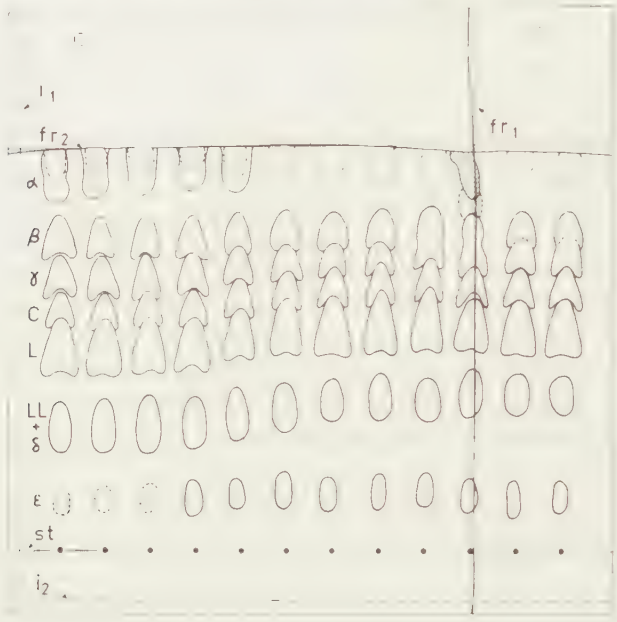

Fig. 9

Fig. 10

Fig. 9. Chromatogram of soybean phosphatides on washed, $SiCl_4$-treated paper. Preceding chromatography to front II, the mobile phase was allowed to ascend to front I perpendicular to the direction of chromatography.

Fig. 10. Chromatogram of the reference mixture on washed, $SiCl_4$-treated paper. Further description as fig. 7.

spot α in unstained form is visible too and lies compressed between the comma-shaped spot and spot β.

Now turning to the positions of the remaining spots, we observe that only the ε spots are situated on a horizontal line. The other rows of spots: LL+δ, L, C, γ and β, lie on lower levels in zone a than in the zones c and d in diagram III of fig. 6. The transition once more takes place in zone b.

The changes in level are less pronounced than in the case of non-impregnated paper. This is a sign that the removal from zone a of the contaminating substances, responsible for the higher position of these spots in zones c and d, has been less effective in impregnated paper than in non-impregnated paper. Possibly this also may account for the fact that in fig. 9 only the spots ε lie on a horizontal level, whereas in fig. 7 also the spots LL and δ were found to lie on horizontal levels.

Now turning to fig. 10, which gives the chromatogram for the reference mixture, we find that the change in level takes place for all phosphatide components present.

Taking the results of the chromatograms fig. 9 and fig. 10 together, we state that on impregnated paper the position of the spots β, γ, C, L, S+LC, LL and δ is influenced by the considered contaminations in the $SiCl_4$-treated paper, whereas only the ε spots are not affected in their position by these contaminations being present or absent.

Acknowledgement

I gratefully acknowledge the skilled technical assistance of Miss J. S. DEN BOER.

0. *Summary*

1. For simultaneously staining on chromatograms spots of amphi-ionic phosphatides (lecithin, kephalin, their lysoproducts and sphingomyelin) red, and spots of acidic substances (inositol phosphatides and others) green, a mixed staining bath (0.006 % Acid Fuchsin, 0.01 % Brilliant Green, 0.2 % $UO_2(NO_3)_2$, and 0.01 N HCl) has formerly been given. The adhering inconveniences of this staining method have been much reduced by altering the composition of the mixed staining bath.

2. The composition of the improved bath (0.02 % Acid Fuchsin, 0.01 % Brilliant Green, 0.04 % $UO_2(NO_3)_2$ and 0.1 N acetic acid) now permits to reduce the staining time to one hour (instead of formerly overnight). The much faster rate of staining is mainly due to the substitution of 0.01 N HCl by 0.1 N acetic acid, while the relatively greater intensity of the green spots resides mainly in the five times smaller $UO_2(NO_3)_2$ concentration. To compensate for the loss in intensity of the red spots, the Acid Fuchsin concentration has been increased.

3. As a result of the substitution of 0.01 N HCl by 0.1 N acetic acid the green background colouration is also strongly increased. It is therefore

necessary to wash out first with a HCl containing washing bath (0.04 % $UO_2(NO_3)_2$ in 0.01 N HCl) during 10–20 minutes. Before drying the chromatogram it is shortly washed with an acetic acid containing washing bath (0.04 % $UO_2(NO_3)_2$ in 0.1 N acetic acid). After drying with a hairdryer, the green spots are now much better visible than with the formerly used staining method. Thus it is no longer necessary to circumline the green spots with a pencil on the still wet chromatogram, as was necessary before.

4. The internal equilibrium and therewith the staining power of the mixed staining bath depends on the temperature. By changing the temperature the shift to the new equilibrium state proceeds but slowly. To obtain reproducible staining results at roomtemperature (20 °C) one should use always a fresh portion of the mixed staining bath from the stock in the refrigerator (5 °C) and stain a fixed time (for instance 1 hour).

5. Some additions (0.5 % phenol or 1 % resorcinol) to the mixed staining bath increases the intensity of the green spots considerably. The addition of phenol or resorcinol should be done just before staining, as sooner or later there appears a flocculation in the mixed staining bath. The washing liquids should contain also ½ % phenol or 1 % resorcinol, lest the intensity of the green spots decreases in the washing process.

As the green background colouration also increases with 0.5 % phenol or 1 % resorcinol a distinct advantage of these additions is only present when paper with reduced acidic contaminations is used.

6. Some methods have been given to reduce the green background colouration of the chromatograms, especially of the green zone about half-way the distance from front to immersion line.

They consist in pretreatments of the paper. Effective are:

a) Bathing the paper for instance over night in 95 % ethanol.
b) Ascension in the paper of some polar liquids (methanol, ethanol, acetone)
c) The best pretreatment found consists in ascension of 86 % n. butanol over night, followed the next day by ascension of cellosolve over night.

7. A survey has been given of attempts to remove by ascension of various liquids in non-impregnated paper contaminations which stain green with Brilliant Green, contaminations which give the Schiff reaction and contaminations which give the ninhydrin reaction.

8. The results show that we indeed must speak in the plural. Especially in the case of "Brilliant Green substances", we certainly have to do with a large number of contaminants. This appears from the fact that with certain liquids a number of green zones appear in the paper background after staining with the mixed staining bath.

9. The complexity of the green zone at the front after ascension with the mobile phase di-isobutylketone – acetic acid – H_2O – 50:25:5 in washed $SiCl_2$ treated paper, comes to the fore when before staining the chromatogram is heated ½ hour at 90 °. We then find after staining four

adjacent zones: a narrow blue zone at the front line followed downwards by a green zone, a violet zone (red on the still wet chromatogram) and a green zone.

10. A method has been given to investigate whether in the chromatographic separation of a phosphatide mixture, contaminations in the paper which move with the green stainable zone about half-way the distance front to immersion line, play a part in the position of the spots. A quadratic paper sheet (24 × 24 cm) is first run with the mobile phase in one direction, then after a week hanging in air the paper sheet is run in a direction perpendicular to the first direction. The phosphatide mixture is applied on twelve starting points on a horizontal line before the second ascension.

11. On the developed chromatogram the twelve spots belonging to the same component may lie either on a horizontal line over the whole width of the chromatogram, or at one side of the chromatogram on a lower level than at the other, the transition from one level to the other taking place in the stainable zone present in the paper after the first ascension.

In the former case the component considered is not influenced in its position by the absence or presence of the contamination mentioned point 10.

In the latter mentioned case an influence is present. The spots are situated higher in the presence of the contamination than in its absence.

12. For further particulars concerning the behaviour of components occuring in chromatograms obtained with a soybean phosphatide preparation and with a mixture of lecithin, cephalin, their lysoproducts, and sphingomyelin, both on non-impregnated paper and on washed $SiCl_4$-treated paper, the text should be consulted.

Department of Medical Chemistry,
University of Leiden

REFERENCES

1. Bungenberg de Jong, H. G. and G. R. van Someren, these Proceedings, Series B, **62**, 150 (1959).
2. Hooghwinkel, G. J. M., J. Th. Hoogeveen, M. J. Lexmond and H. G. Bungenberg de Jong, these Proceedings, Series B, **62**, 222 (1959).
3. Bungenberg de Jong, H. G. and J. Th. Hoogeveen, these Proceedings, Series B, **63**, 190 (1960).
4. ——— and ———. I. These Proceedings, Series B, **63**, 190 (1960).
   II. „ „ „ „ **63**, 228 (1960).
   III. „ „ „ „ **63**, 383 (1960).
   IV. „ „ „ „ **64**, 1 (1961).
   V. „ „ „ „ **64**, 167 (1961).
5. ——— and ———, these Proceedings, Series B, **63**, 1 (1960).

HYDROSOLS OF TOTAL EGG-PHOSPHATIDES. IIa

A STUDY OF THE SURPLUS NEGATIVE CHARGE IN HYDROSOLS

BY

H. G. BUNGENBERG DE JONG AND J. TH. HOOGEVEEN

(Communicated at the meeting of April 29, 1961)

1. Introduction

In Part I of this series [1] we have described the preparation of high and low disperse sols of purified total egg phosphatides, starting from a commercial phosphatide preparation (Lecithin ex ovo puriss. from Merck). High and low disperse sols showed a striking difference towards low concentrations of uranyl salts, or towards low concentrations of long chain alkyl trimethylammonium salts.

High disperse sols: Strong turbidity, disappearing within a few seconds; no flocculation

Low disperse sols: Permanent flocculation zone.

It has been suggested that the primary cause for both phenomena is the same, namely that the cation of the salts mentioned is bound by a negative amphipatic substance, which is present in the purified egg-phosphatides in small concentrations. In Part I a method has been given to determine this surplus negative charge in high disperse sols, making use of the above mentioned transient turbidity. For a systematic study of this surplus negative charge (origin, pH dependence, factors influencing it; changes with time) this method is not suitable. In the present communication we will therefore use lower disperse sols and study the resulting turbidity or flocculation zones. The surplus negative charge, as will be shown, is present already in the commercial phosphatide preparation. The negative charge thus is not only the result of the purification and (or) the procedures followed in preparing the sols, although it may be increased thereby. Experiments will show how to avoid this increase. We extended the investigation to total egg-phosphatides prepared from hen's eggs. Here too the sols showed a surplus negative charge. This charge showed a different dependence on the pH and a slower increase with time under the influence of the atmospheric air. These differences will be reviewed in the final discussion of the experimental results.

2. Determination of the LTAB number, a measure for surplus negative charge

a) *Choice of the salt appropriate to determine the surplus negative charge*

Uranyl ions and long chain alkyl trimethylammonium ions are almost quantitatively bound to the surplus negative charge. However, uranyl salts are only applicable in acid medium. At pH = 7, for instance the uranyl ion hydrolyses to uranate. Since we want to determine the surplus negative charge also at about pH 7, long chain alkyl trimethylammonium-salts are indicated for our purpose. We had the choice between CTAB (= cetyl trimethylammonium bromide) and LTAB (= linoleyl trimethylammonium bromide). The former has the disadvantage that it slowly separates from watery solutions kept at room temperature; this does not occur with LTAB. In most experiments we followed the change of the surplus negative charge over a period of several days and therefore the choice fell on LTAB. Solutions of LTAB proved to be perfectly stable. Changes of the surplus negative charge with time proved not to be due to alterations of the LTAB.

b) *Quantitative binding of the LTA-cation at maximum flocculation*

As a first example we give the determination of the surplus negative charge of the commercial preparation "Lecithin ex ovo puriss. (Merck)", which had not been purified any further. We started from a 3 % solution in 30 vol. % tertiary butanol (30 ml tertiary butanol + 70 ml H_2O). With the aid of the apparatus described in Part I of this series (see there, fig. 7), 30 ml of this 3 % solution (3 gr in 30 ml) was injected while rigorously stirring into 60 ml of H_2O.

The primary sol obtained has a more or less turbid appearance. It contains 1 % phosphatides and 10 vol % of tertiary butanol. This primary sol is used as such for the determination of the surplus negative charge. In test tubes two series of mixtures were prepared of the composition:

A) a ml LTAB 2 mmol/l + (1 − a) ml H_2O

B) a ml LTAB 2 mmol/l + (1 − a) ml H_2O + 9 ml 6.67 vol % tertiary butanol.

In these mixtures a was increased by steps of 0.05 ml. We added 1 % primary sol to each tube 2 ml of the abovementioned, and mixed them. In the two series the end concentration of the tertiary butanol is the same, namely 6.67 vol %.

As the end volumes are different (A = 3 ml; B = 12 ml) the end concentration of the phosphatides are in series A (0.667 %) four times higher than in series B (0.167 %). It is interesting that nevertheless the flocculation zone (which had practically not altered in 2 hours) is found in both series at practically the same values of a. Compare columns 2 and 4 of Table I. To estimate which value of a corresponds to the flocculation optimum, we may use various criteria, e.g. the intensity of flocculation.

the eventual difference in opalescence of the "serum" remaining, the rapidity with which after shaking flocules become visible with the naked eye. In column 2 the appearance of the flocculated system at a = 0.35 and a = 0.40 was the same, and also the opalescence of the sera, but after shaking the flocules became visible faster in the test tube with a = 0.4, than in that with a = 0.35. Hence we estimated here for a corresponding to optimal flocculation a value somewhat nearer to 0.4 than to 0.35, namely a = 0.38. In column 4, where only one tube shows flocculation (strong with nearly clear serum) we estimated a of this tube (0.40) to represent optimal flocculation.

TABLE 1

Flocculation zones with LTAB in the series A and B before and after acidifying

a	Serie A		Serie B	
ml LTAB 2 m mol/l	end volume 3 ml	after acidifying, end volume 4 ml	end volume 12 ml	after acidifying, end volume 16 ml
0	opalescent	opalescent	opalescent	opalescent
0.05	,,	,,	,,	,,
0.1	,,	,,	,,	,,
0.15	,,	,,	,,	,,
0.2	,,	++++	,,	+++
0.25	,,	++++	,,	+++(+)
0.3	,,	++	,,	opalescent
0.35	+++	turbid	,,	,,
0.4	+++(+)	weakly turbid	++++	,,
0.45	opalescent	opalescent	turbid	,,
0.5	,,	,,	opalescent	,,
0.55	,,	,,	,,	,,
0.6	,,	,,	,,	,,
estimated a at flocculation optimum	0.38	0.23	0.40	0.23

Hereafter we added 1 ml 0.1 N acetic acid to each tube of series A and 4 ml 0.1 N acetic acid to each tube of series B and shook the mixtures. The concentration of acetic acid is now the same in both series, namely 0.025 N. The phosphatide concentrations however have decreased to 3/4 the original values. The end volumes are now 4 ml in series A, and 16 ml in series B. We let the tubes stand 2 further hours, after which time we obtained the readings recorded in columns 3 and 5 of Table I respectively. We here estimated 0.23 and 0.23 respectively as a values corresponding to optimal flocculation.

The experimental results may now be used to calculate the end-concentrations of LTAB corresponding to optimal flocculation and the phospha-

tide end concentration in series A and B before and after addition of acetic acid. We then obtain the values given in the next survey:

	not acidified			in the presence of 0.025 N acetic acid		
	a (opt. flocc.)	phosphatide %	LTAB m mol/l	a (opt. flocc.)	phosphatide %	LTAB m mol/l
Series A	0.38	0.667	0.253	0.23	0.500	0.115
Series B	0.40	0.167	0.067	0.23	0.125	0.029

In fig. 1 the values of column 4 have been plotted against those in column 3, similarly the values of column 7 against those in column 6. It appears that within the errors of the estimation the two points belonging to the non-acidified series A and B lie on a straight line through the origin. The same applies for the two points belonging to the acidified series, only the slope of the straight line is smaller. We may draw two important conclusions:

a) At optimal flocculation the LTA cation present is practically completely bound to the surplus negative charge present in the phosphatide preparation.

b) The surplus negative charge is a function of the pH.

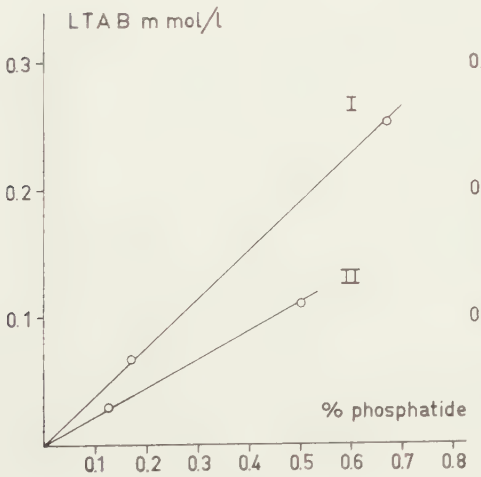

Fig. 1

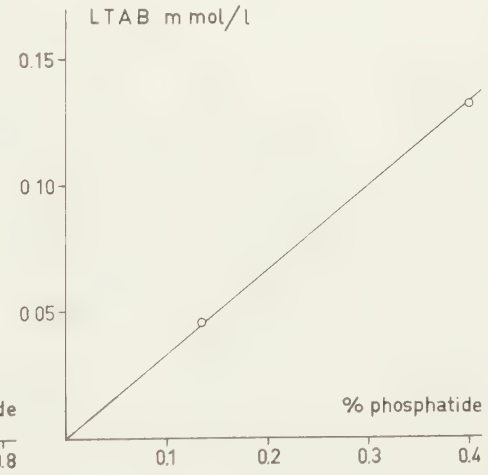

Fig. 2

Figs. 1 and 2. The LTAB concentration corresponding to maximum flocculation as a function of the phosphatide concentration. Explanation see Text.

c) *Equivalent binding of the LTA cation in buffered systems*

The above point b) shows that is desirable to regulate the pH when determining the surplus negative charge. The pH of sols may vary according to the purification applied to the commercial phosphatide preparation. Sols obtained from total egg-phosphatides after purifying by 3 times

precipitation of the etherial solution with acetone, followed by extraction with CCl_4 from the solution in 60 vol % ethanol, (for details see Part I of this series) have a higher pH than when the extraction procedure is omitted.

Bromophenol Blue added to the former sols gives a blue colour; if added to the latter sol (or to a sol which is made from the phosphatide preparation without previous purification with acetone from ethereal solution) it gives a yellowish green colour.

In the experiments of this paragraph we used 3 % sols prepared from the commercial preparation after threefold precipitation with acetone from the ethereal solution. With the aid of the apparatus given in part I of this series, 30 ml of a 3 % solution of this purified phosphatide in 30 vol % tertiary butanol was injected while stirring into 60 ml H_2O. The sol obtained was strongly opalescent (because the extraction procedure with CCl_4 from 60 % ethanol following the purification from ethereal solution with acetone, had been intentionally omitted) and showed a flocculation zone with LTAB (which was aimed at). In preliminary experiments it was found that NaCl when present in the flocculation series at an end concentration of 20 m mol l led to a broadening of the flocculation zone with LTAB, possibly also to a slight shift of the flocculation zone towards higher LTAB concentrations.

As we intended to use a Na-acetate – acetic acid buffer, and since Na acetate belongs to the same salt type (monovalent cation-monovalent anion) as NaCl, it may be expected that the buffers will in principle have the same effect as NaCl. As both influences are undesirable, we must take care that in the end mixtures of the flocculation series the concentration of Na-acetate is lower than 20 m mol l, but still high enough to ensure a safe pH regulation.

In the following experiment the Na-acetate concentration is 10 m mol l (in later experiments in this communication still lower, namely 5 m mol l).

Two series of mixtures were prepared in test tubes:

Series A: a ml LTAB 2 m mol/l + (2 − a)ml H_2O + 1 ml acetate buffer (1 : 1).

Series B: a ml LTAB 2 m mol/l + (10 − a)ml H_2O + 3 ml acetate buffer (1 : 1) in which the buffer used consisted of 1 volume Na acetate 0.1 N + 1 volume acetic acid 0.1 N. Then to each tube 2 ml of the above mentioned 1 % phosphatide sol was added, after which addition the mixtures in series A and B have an end volume of 5 and 15 ml respectively, while the end concentration of the Na-acetate is in the two series the same, namely

	a	LTAB m mol l^{-1}	phosphatide %
series A	0.33	0.132	0.40
series B	0.34	0.045	0.133

10 m mol/l. The reader will be spared an extensive Table analogous to Table I, but in the following survey the values of a corresponding to optimal flocculation and the end concentrations of LTAB and phosphatide are given.

In fig. 2 the LTAB concentration has been plotted against the phosphatide concentration. We see that, as in fig. 1, the points lie on a straight line through the origin. We conclude:

a) In the presence of acetate buffer of small concentration, at optimal flocculation the LTA cation is still practically completely bound to the surplus negative charge in the phosphatide preparation.

b) The purification applied to the commercial preparation (3 times precipitation with acetone from the ethereal solution) has not removed the surplus negative charge.

Point a) opens the possibility of investigating the influence of the pH on the surplus negative charge, which problem is dealt with further in section 6 below.

Point b) makes it improbable that the surplus negative charge is simply due to a contamination of the phosphatide preparation with some fatty acid.

d) *Amount of phosphatide to which is associated one equivalent surplus negative charge*

A phosphatide preparation containing only phosphatides with one positively charged and one negatively charged group (lecithin, kephalin, their lyso products and sphingomyelin) will not have a surplus negative charge in the pH range of 4.5 — 6. The surplus negative charge present at for instance pH 4.7, must be due to one or more acidic substances on the nature of which we will not speculate here. From the determinations as in paragraphs b and c we may get an idea how much of these acidic substances are present in the preparation. As in fig. 2 the straight line through the two points proceeds through the origin, the LTA cation at optimum flocculation is not free but practically completely bound to the surplus negative charge. We read from the figure that at a phosphatide concentration of 0.4 %, that is 4 g/l, the necessary "concentration" of LTAB amounts to 0.133 m.equiv/l LTAB. From this follows that one equivalent surplus negative charge is associated with about 30.000 g phosphatide. Taking arbitrarily the average molecular weight of the phosphatides equal to 750, and the equivalent weight of the substance(s) carrying this negative charge also at 750, then one equivalent of the latter will be present on 39 phosphatide molecules. This gives a content of 2.5 equivalent %. This value is not very sensitive to the equivalent weight assumed. When we take the latter two times the molecular weight of the phosphatides, one obtains 1/39 = 2,56 %. When on the contrary the equivalent weight is much smaller, for instance as a limit zero, one would obtain 1/41 = 2.44 %. Similar calculations may be made from the slope of the two straight lines in fig. 1. At 0.5 % phosphatide concentration we read here of 0.191 (I) and 0.112 (II) mmol/l LTAB respectively. From these data follows that one equivalent negative charge is associated with about 27.000 g (case I) and 45.000 g (case II) of the phosphatide preparation respectively. Taking arbitrarily the mean molecular weight of the phosphatides 750 and the equivalent weight of the substance(s) carrying the surplus negative charge also 750, one obtains a content

of about 2.8 (case I) and 1.7 (case II) equivalent percents of the phosphatide preparation.

c) *Simplified method for the determination of the surplus negative charge. The LTAB number*

The results in paragraphs b and c show that at the flocculation optimum the LTA cation present is practically completely bound to the surplus negative charge in the phosphatide sol. This allows the simplification that the surplus negative charge does not have to be found from two or more flocculation series with different phosphatide end concentrations. One flocculation series already suffices.

In the following sections, in which the influence of various factors, time included, will be investigated, the mixtures in the flocculation series have generally the composition:

a ml LTAB 1 m mol/l (with steps of 0.05 or 0.1 ml)
(7 − a) ml H_2O
1 ml buffer
2 ml sol 1 %

We now will call the flocculation optimum obtained at a value of a divided by 2 the LTAB number.

$$\text{LTAB number} = \frac{a_{opt}}{2}.$$

With the above choice of the composition of the mixtures half the numerical value of a_{opt} at the flocculation optimum, at once gives the surplus negative charge expressed in simple units. In a_{opt} ml LTAB 1 m mol l are present $a_{opt} \cdot 10^{-6}$ m.eq. LTA cations and these are bound to the negative charge associated with the 2.10^{-2} g. phosphatide present in the 2 ml 1 % phosphatide sol. The LTAB number thus gives the negative charge associated with one gram of the phosphatide expressed in 10^{-4} equivalents.

3) Shift of the flocculation zone with time, and resulting difficulties of estimating the LTAB number

In the experiments on the quantitative binding of the LTA-cation in section 2 (see figs. 1 and 2) the flocculation optimum was read off after a few hours standing. In these examples the position of the flocculation optimum remained constant over a period of several hours. It has appeared that the flocculation zone begins to shift towards higher LTAB concentrations sooner or later. At the same time the flocculation zone becomes broader (compare scheme in fig. 3). This already leads to larger errors in estimating the position of the optimal flocculation, several tubes having the same aspect. One may then take as flocculation optimum the mean LTAB value of a_{opt} of the first and last tube showing the same aspect. When one shakes the tubes (10 times after closing with a rubber stopper)

and then waits until the flocculation just becomes visible, one finds in general a higher value of a_{opt}, as could be expected from the time elapsed between the two readings. This is due to the fact that the spontaneous peptisation at rest of a flocculated system to an opalescent sol is but a slow proces. Hence the a_{opt} value of a flocculation zone lies higher in

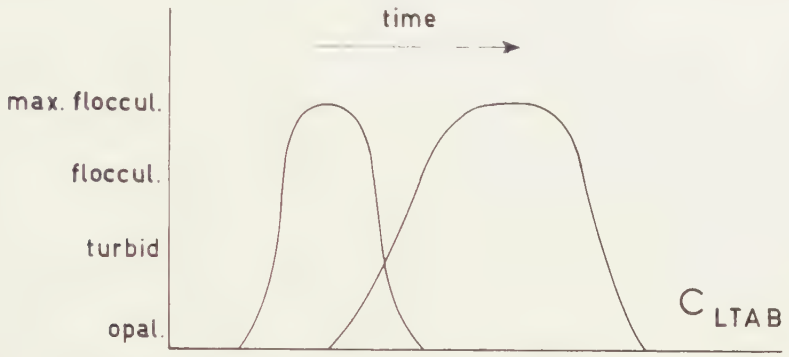

Fig. 3. Scheme showing the shift of the flocculation zone with time in the direction of higher LTAB concentrations. Explanation see Text.

reality than the mean of the a values of the first and last tube showing the same aspect. When over a period of several days one estimates once or twice each day a_{opt}, first at rest and then shortly after shaking (1 – 2 hours), and plots the readings against time, one does not, in general, obtain a smooth curve, but a more or less zig-zag curve, as shown by the figures following below. Thus in studying the influence of factors on the rate of increase of the surplus negative charge, weak influences will disappear in the estimation errors and only strong influences come to the fore.

4) Measures to be taken in the preparation of hydrosols to avoid an increase of the surplus negative charge

The sol used in section 2, paragraph c showed at pH 4.65 a rather high surplus negative charge (LTAB number 0.33 and 0.34 respectively). To prepare the sol we started from the commercial Merck preparation which was three times precipitated with acetone from ethereal solution. The precipitate was dissolved in ether, poured in a large cylindrical flask of known weight. By means of the apparatus given in Part I of this series, the ether was evaporated by blowing hot air into the flask, thus leaving a thin layer of phosphatides on the wall of the flask. We continued to blow hot air into the flask, until constant weight was reached. We thus knew the weight of the dry phosphatides on the glass wall and added 30 vol % tertiary butanol to obtain a 3 % solution. This solution then served to prepare the primary sol. It was found that sols made from the Merck preparation without the above purification, thus by simply starting from a 3 % solution of the crude preparation in 30 vol % tertiary butanol,

showed with LTAB a much lower surplus negative charge (LTAB number in the order of 0.1).

Thus it was clear that in the above "purification" we have somehow increased the surplus negative charge. It seems reasonable to suppose that this increase took place during the time a current of hot air was blown along the thin phosphatide layer on the glass wall until constant weight was reached. To investigate this further a 1 % sol of purified phosphatides was made whereby we avoided the above hot air treatment of the purified phosphatides to constant weight. For this purpose we started from a known quantity of the Merck phosphatide preparation and collected all supernatant liquids of the precipitations, evaporated them to constant weight, whereafter by subtraction the weight of purified phosphatide is known.

The precipitate was then dissolved in the calculated amount of 45.5 % tertiary butanol to obtain a 4 % solution. Three 1 % sols A, B and C were made from this solution as follows:

Sol A: To 15 ml H_2O was pipetted while stirring 5 ml of the 4 % solution.

Sol B: 5 ml of the 4 % solution was evaporated on the wall of a rotating large cylindrical flask, while air of 48° C was blown into the flask. Soon the phosphatide layer was formed on the wall and thereafter we continued to blow air of 48° C along it for one hour. After cooling the flask to room temperature, we added 4.8 ml 45.5 % tertiary butanol, in which the phosphatide was dissolved by rolling the flask. We then added 15 ml H_2O with a pipette.

Sol C: The same as sol B, with the only difference that the phosphatide layer on the wall of the flask was exposed during one hour to a stream of air of 100° C.

We then determined the negative surplus charge with each of the three sols, whereby we not only restricted ourselves to the flocculation zone a short time after starting the experiment, but followed also the shift of the flocculation zone during some 60 hours.

The results, compare fig. 4, confirm the supposition mentioned above, that exposure of the nearly dry phosphatide layer on the glass wall to a stream of hot air increases the initial value of the surplus negative charge. With sol A, we find a LTAB number of the same order as with sols of non purified Merck phosphatides. With sol B (exposure 1 hour to air of 48° C) the initial value is higher, and with sol C (exposure 1 hour to air of 100° C) the initial value is still higher. We further see that for all three sols the surplus negative charges increase with time, whereby the curves show the zig-zag character, discussed in section 3. The curve for sol C could not be followed further as given in the figure, as the total number of tubes used in this experiment was too small.

The method given above to obtain a solution of known concentration of the purified egg-phosphatides, without drying the phosphatide layer on the wall of the flask with a current of hot air to constant weight, is relatively inconvenient. A simpler method, consists of dissolving the purified phosphatide in ether, and evaporating the ether spontaneously (without blowing) in a rotating flask untill the solid layer on the wall has formed, then to dissolve the phosphatide in a tertiary butanol – water mixture, to obtain a concentrated solution. We then determine the dry weight on a sample of this solution, from which follows how much of the tertiary butanol – water mixture must be added to the remaining solution to obtain the desired phosphatide concentration.

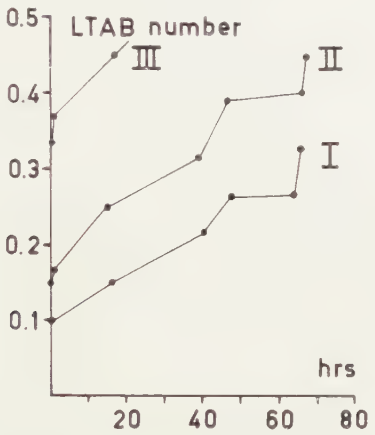

Fig. 4. Influence of exposure of the phosphatides in a thin layer to warm air on the surplus negative charge of sols prepared from these phosphatides, and on its change with time. I blank (no exposure); II 1 hour exposure to air of 48°; III 1 hour exposure to air of 100°. Explanation see Text.

The results of this section show that one should avoid exposing a dry phosphatide layer to a current of hot air. The blowing technique described in Part I of this series [1] nevertheless remains a convenient and not dangerous method to concentrate a solution or to remove tertiary butanol from a primary sol. One may conveniently, without danger, blow hot air into the rotating flask. As long as solvent is present the solution remains relatively cold by evaporation (for instance 25°) although we used air of 80 – 100° C.

5) Influence of the temperature on the increase of the surplus negative charge with time

In the experiments in this section we used a 1 % sol from the non-purified Merck preparation. Four identical series of tubes with increasing LTAB concentration were made, two of which were buffered (acetate buffer) to pH 4.65, the two others with phosphate buffer to pH 6.95. The shift of the flocculation zone was followed at 17° C during about

140 hrs. After 66 hrs. one series of pH 4.65 and after 42 hr one series of pH 6.95 was put in the refrigerator (about 5° C) and was left there to the end of the experiment.

The results — compare fig. 5 — show:

1) At pH 6.95 a higher initial surplus negative charge is found than at pH 4.65 (compare already in fig. 1 the larger slope of line I compared to the slope of line II).
2) At 17° C the rate of increase of the surplus negative charge is greater at pH 6.95 than at pH 4.65 (compare curves I).
3) The increase of the surplus negative charge is much slower at 5° C. Compare curves II with curves I.

The results suggest that the increase of the surplus negative charge with time is due to a chemical reaction.

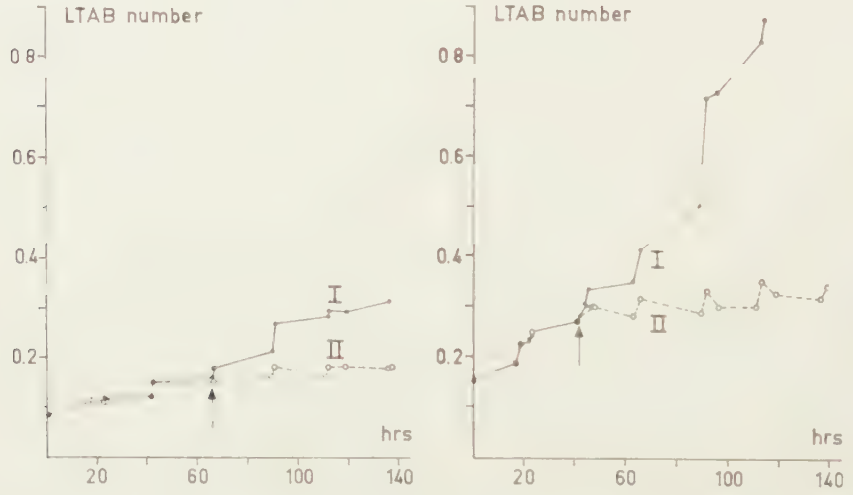

Fig. 5. Influence of the temperature on the rate of the increase of the surplus negative charge with time at two pH values. (Lefthand graph at pH 4,65, righthand graph at pH 6,95). Explanation see Text.

6) Influence of the pH on the initial surplus negative charge

This influence has been investigated a number of times. The difficulty here is that with the increase of the pH the rate of increase of the surplus negative charge with time also strongly increases. We give therefore as example the results of a series which was performed at 5 C. The sol used was made from the non-purified Merck preparation. With the aid of different acetate and phosphate buffers flocculations series were made at pH 3.23; 3.69; 4.71; 5.71; 6.09; 6.42; 6.72; 7.02; and 7.48 (the pH was measured of the final-mixtures containing the sol). The readings of the optimal flocculation were taken 3 hours later and have been plotted in fig. 6. The S-shaped curve, when extrapolated downwards to the left,

meets the abscissa at a pH somewhat lower than 3. We know from former work that about here (pH 2.85) the apparent isoelectric point of the egg-phosphatide preparation of Merck is situated.

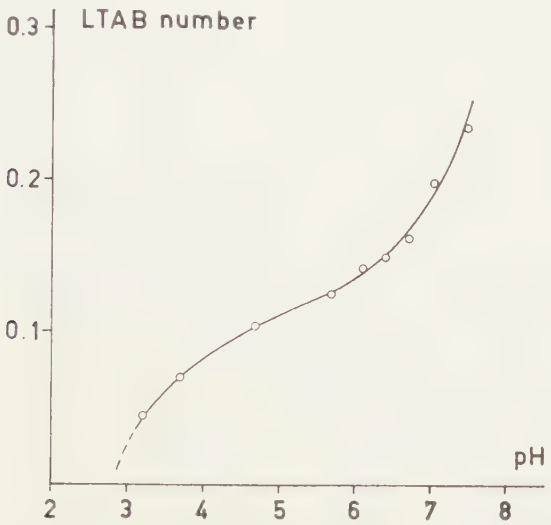

Fig. 6. Influence of the pH on the LTAB number at 5° (readings taken after 3 hours). Explanation see Text.

With other sols made of purified or non-purified egg-phosphatides of the Merck preparation we found at 20° C similarly shaped curves as in fig. 6; the extrapolated starting point on the abscissa lying at about pH 3, though the absolute values of ordinate values varied. In all these cases the surplus negative charge at pH 6.85 is higher than at pH 4.65.

7) Influence of ether-peroxides and of H_2O_2

As in the purification of egg-phosphatides ether is used as solvent, and because this may contain peroxides, we have compared sols of non-purified egg-phosphatides containing ether-peroxides (by dissolving 4 gr egg-phosphatides in 125 ml peroxide-containing ether and evaporation of the ether) with sols from egg-phosphatides not so treated. Similar experiments have been performed as in fig. 5. Within the experimental errors we could not find a difference in the initial surplus negative charges, nor in the rate of its increase with time.

In further experiments it was found that in the presence of 0.06 % H_2O_2, the rate of the increase of the surplus negative charge with time is within the experimental errors the same as without H_2O_2. The latter result is important because H_2O_2 in this concentration prevents the growth of micro-organisms. It is therefore very improbable that the chemical reaction responsible for the increase of the surplus negative charge with time as supposed at the end of section 5 is due to micro-organisms.

8) Influence of oleic acid, n. heptylic acid and n. heptyl aldehyde

The chemical changes occurring with time in the hydrosol might be a slow hydrolysis, setting free mainly an unsaturated fatty acid (for instance oleic acid), or an oxydative attack on an unsaturated fatty acid esterified in the phosphatide molecule.

This induced us to investigate the influence of oleic acid, n. heptylic acid (= heptanoic acid) and n. heptylic aldehyde (= heptanal) at pH 4.65 and pH 6.85.

The abovementioned additions to an extent of 4 or 8 mol percent of the egg-phosphatide, the mean molecular weight of the latter arbitrarily taking 750, were dissolved in the tertiary butanol-H_2O solution of the egg-phosphatides (Merck preparation, not purified) which served for preparing the sols.

Oleic acid

The experimental results at pH 4.65 are given in the left graph of fig. 7. It is seen that at this pH oleic acid has no influence on the initial surplus negative charge and that we cannot conclue with certainty to an influence on the rate of the increase of the surplus negative charge with time.

Quite different results are found at pH 6.85 (compare right graph in fig. 7). The initial value of the surplus negative charge is strongly

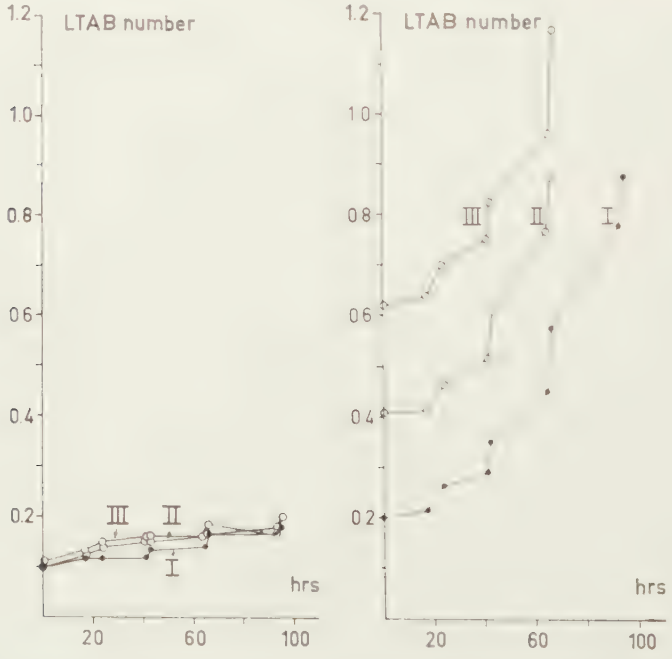

Fig. 7. Influence of oleic acid on the surplus negative charge and on its change with time. Lefthand graph at pH 4.65, righthand graph at pH 6.85. Curve I = blank; Curve II and III with 4 and 8 molecule percents oleic acid added. Explanation see Text.

increased by oleic acid, proportionally to the amount added. The increase amounts to 0.21 at 4 mol % and 0.42 at 8 %, thus about 0.05 for 1 mol % oleic acid. The surplus negative charge (LTAB number, see section 2) being expressed in 10^{-4} equivalents per gram phosphatide, we must multiply 0.05 with 800 (the arbitrarily taken mean molecular weight of egg-phosphatides) to obtain the equivalents surplus negative charge per mol phosphatide. We then obtain 40×10^{-4} or 0.4×10^{-2} extra surplus negative charge for 1 mol per cent oleic acid, that is 10^{-2} mol oleic acid per mol phosphatide.

Hence at pH 6.85 the oleic acid present in the phosphatide sol (taken up in its micelle) is not wholly dissociated, but only less than half. As at lowering the pH the dissociation of the oleic acid will be diminished further, it is not surprising that at the two units lower pH in fig. 6 (left graph) the addition of oleic acid did not increase the initial surplus negative charge.

Heptylic acid

We have investigated in a similar way as above for oleic acid the influence of 4 and 8 mol % heptylic acid at pH 4.65 and pH 6.85. It has been found that at both pH no influence could be observed on the initial surplus negative charge nor on the rate of increase of the surplus negative charge with time. The difference with oleic acid (in a figure analogous to the right graph in fig. 7, the three curves coincide within the experimental errors) very probably results from the relative short hydrocarbon chain of the heptylic acid. Heptylic acid will probably not be absorbed into the phosphatide micelle, whereas oleic acid will be taken up completely.

Heptylaldehyde

In a similar manner as above with oleic acid and heptylic acid, the influence of 8 mol % heptyl aldehyde has been investigated. Compare fig. 8. It appears that at both pHs heptyl aldehyde has no influence on the initial surplus negative charge. It increases however the rate of increase of the surplus negative charge with time. This increase, distinctly present at pH 4.65, is much greater at pH 6.85.

Discussion

The results in this section point in the direction of an oxydative attack of the phosphatides rather than in the direction of hydrolysis. Long chain fatty acids set free by hydrolysis cannot explain the observed increase of the surplus negative charge at pH 4.65.

It is known that aldehydes are formed when phosphatides are exposed to air (Schiff-reaction). It is interesting that heptyl aldehyde increases the rate of the production of surplus negative charge. This cannot reside in a dismutation of the aldehyde into a fatty acid and an alcohol (2 heptyl-

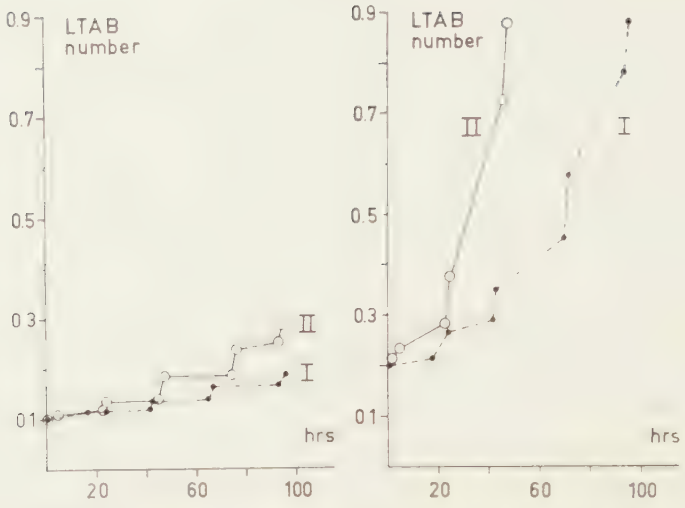

Fig. 8. Influence of heptylaldehyde on the surplus negative charge and on its change with time. Lefthand graph at pH 4,65, righthand graph at pH 6,85. Curve I = blank; Curve II with 8 molecule percents added heptylaldehyde. Explanation see Text.

aldehyde - 1 heptylic acid - 1 heptanol) because heptylic acid appeared to exert no influence.

Hence the aldehyd itself must be responsible for the extra production of surplus negative charge in egg-phosphatides. A discussion how this might be possible will be postponed until we have demonstrated that the increase of the surplus negative charge with time is indeed the result of oxydation by air.

(To be continued)

HYDROSOLS OF TOTAL EGG-PHOSPHATIDES. IIB

A STUDY OF THE SURPLUS NEGATIVE CHARGE IN HYDROSOLS

BY

H. G. BUNGENBERG DE JONG AND J. TH. HOOGEVEEN

(Communicated at the meeting of April 29, 1961)

9) CHANGE OF THE SURPLUS NEGATIVE CHARGE IN SOLS WITH AND WITHOUT CONTACT WITH AIR

a) *Methods*

The sol used in the present section has been made from purified egg phosphatides, starting from the Merck preparation, exactly following the purification procedure described in section 7 of Part I of this Series [1]. In principle the purification consists in precipitating the ethereal solution 3 times with acetone, followed by removal of still present amino acids, by shaking the CCl_4 solution with 60 vol. % ethanol 95 %. This time we took the measures described in section 4 of the present communication to prevent that in preparing the 4 % solution in 40 % tertiary butanol — H_2O, the surplus negative charge is increased.

The sol employed was made by pipetting while stirring 1 volume of the above mentioned 4 % solution into 3 volumes of H_2O. The resulting 1 % sol thus contains 10 % tertiary butanol. As according to the recipe for the mixtures in the flocculation series (section 2, paragraph e) 1 volume of sol is added to 9 volume buffered LTAB solution, the end concentration of tertiary butanol in the mixtures is 1 %. As for our present purpose we need two identical series of mixtures, the volume of each mixture was taken twice as large as usual.

We first made in open tubes the 16 ml mixtures of LTAB, buffer and H_2O, then added the 4 ml sol. After closing with a rubber stopper and mixing the contents, we filled from the open tubes a corresponding series of short tubes, of 8 ml capacity and provided with a ground glass stopper, to the brim. We then closed the tubes with the glass stoppers, the excess of the sol mixture flowing over. We must take care in closing the tubes, that no air remains below the stopper [1]). After filling the tubes, there remain some 10 ml in the original tubes. In this way we obtain two series of tubes, one in open tubes and one in closed tubes. Both series have been placed in a dark room at 20° and the shift of the flocculation-zone has been followed some 12 days, whereby a number of days before ending the observation the stoppers of the closed tubes were removed.

b) *Flocculation-series at* pH 4.65

The results have been plotted in fig. 9. We see that the initial surplus negative charge for both series lies at about 0.1. In the first two days

---

[1]) Before filling the 8 ml tubes, they were each provided with a small glass-sphere. By agitating the filled and closed tube one may thus mix the contents, in order to conclude in which tube the flocculation become visible first, that is, where the flocculation optimum is situated.

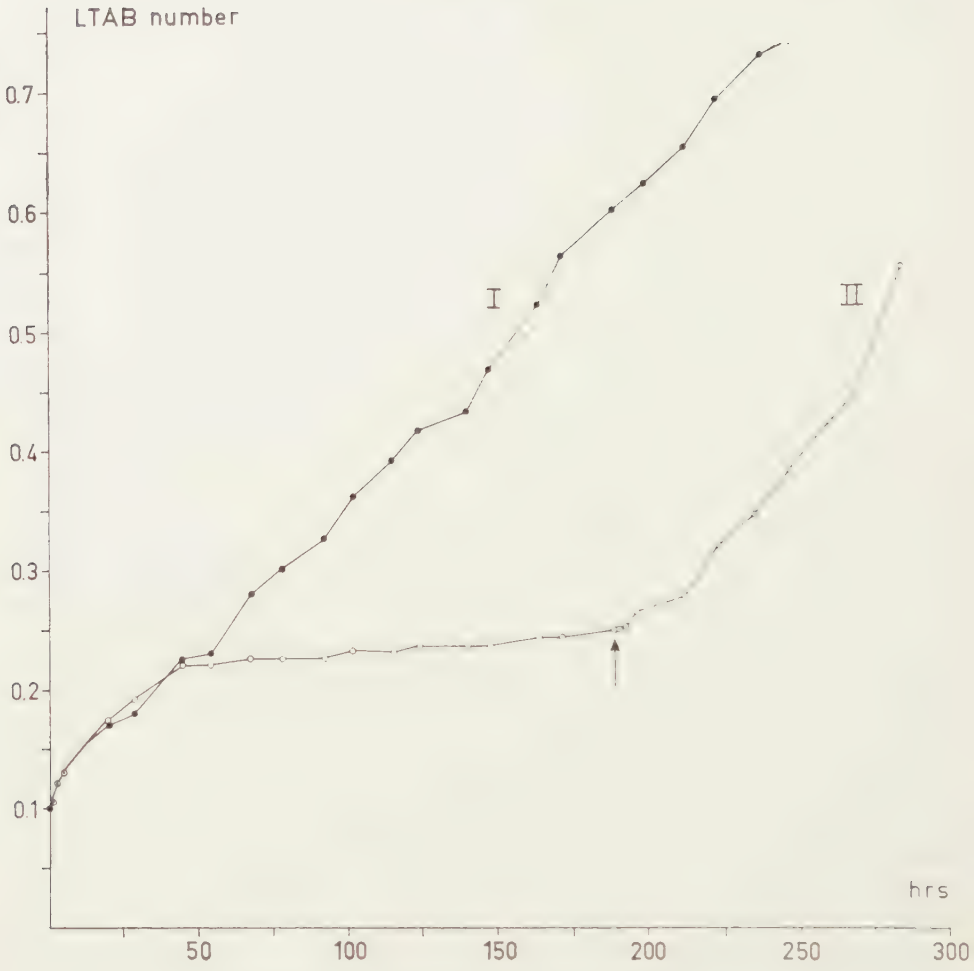

Fig. 9. Change of the LTAB number with time at pH 4,65. I in open tubes; II in closed tubes, the stoppers being removed 190 hours after the beginning of the experiment. (see arrow). Explanation see Text.

there is no difference between the shift of the surplus negative charge in the closed tubes and the open tubes. Thereafter in the case of the closed tubes, the increase of the surplus negative charge with time more or less stops (or become very small). In the open tubes the increase of the surplus negative charge with time continues to proceed further. The stoppers of the series of closed tubes were removed 190 hours after the beginning of the experiment, indicated in the figure by an arrow. We observe that from now on a renewed increase of the surplus negative charge with time sets in.

c) *Flocculation series at* pH 6.85

The results — compare fig. 10 — are quite analogous to those at

pH 4.65. Once more the increase of the surplus negative charge slows down considerably after a few days in the closed tubes, whereas in the open tubes the increase with time continues.

After 240 hours the stoppers were removed from the closed tubes (see arrow), whereafter a strong increase of the surplus negative charge with time sets in.

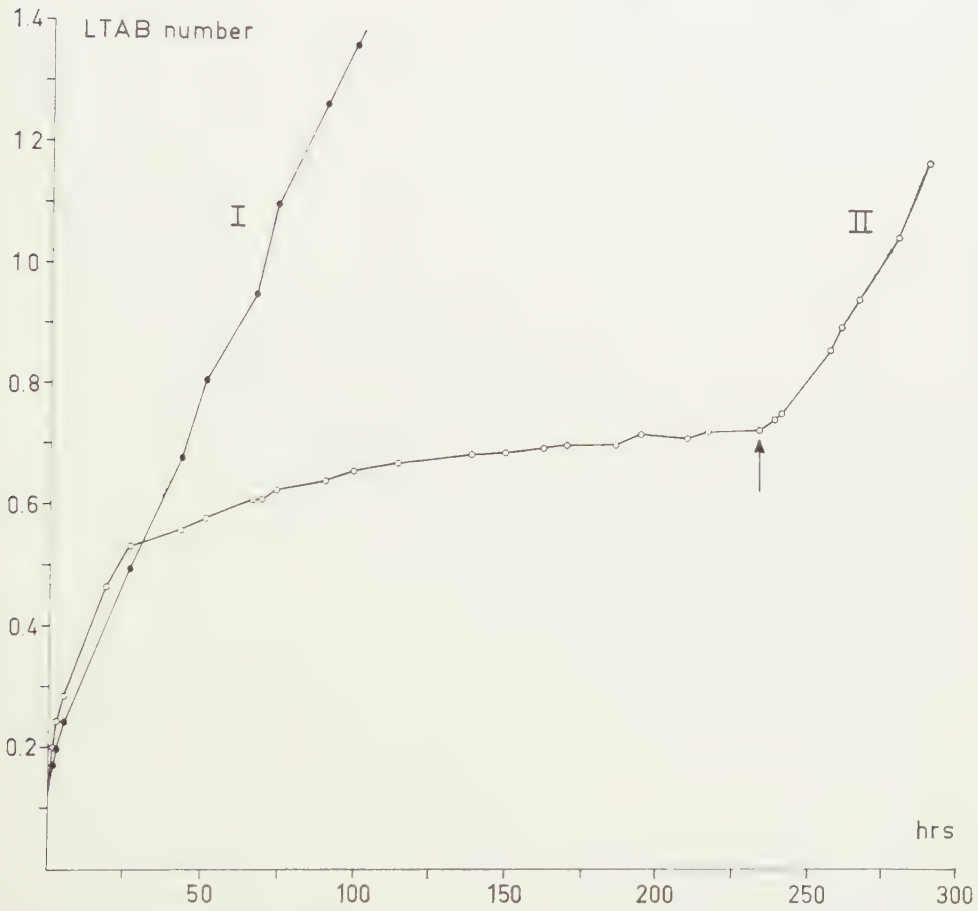

Fig. 10. Change of the LTAB number with time at pH 6,85. I in open tubes, II in closed tubes, the stoppers being removed 240 hours after the beginning of the experiment. (see arrow). Explanation see Text.

d) *Discussion*

The above experiments show clearly that for a continued increase of the surplus negative charge with time the sol must remain in contact with air. This result makes it already extremely probable that increase of the surplus negative charge with time is due to oxydation. That in the closed tubes the increase of the surplus negative charge nevertheless at first takes place in the same way as in the open tubes, must then be due to the sol mixtures being originally saturated with air. The oxydation can

thus proceed at the cost of the dissolved oxygen. The increase of the surplus negative charge with time will then stop or considerably slow down, when the dissolved oxygen has been practically used up. After removing the stoppers, oxygen again dissolves in the sol mixtures and the oxydation can start anew, resulting in a renewed production of surplus negative charge.

10) Experiments in closed tubes with normal as well as with much reduced concentration of dissolved oxygen

In these experiments two series of tubes of larger capacity (50 ml) have been used. In each series we added per tube a ml LTAB 1 m mol/l, increasing a in steps of 0.1 ml, and (2.4 − a) ml H_2O. To each tube was further added 1 ml acetate buffer (a mixture 1:1 of 1 m. Na-acetate and 1 m. acetic acid) and after this 2 ml 2 % phosphatide sol, freed from tertiary butanol [1]).

One series of tubes was then filled with ordinary distilled water (saturated with air), the other with distilled water from which about 97 % of the dissolved oxygen had been removed by a current of N_2 (oxygen was estimated by the method of L. W. Winkler [2])). The results are given in fig. 11. Both curves start from an initial LTAB number [3]) of about 0.08.

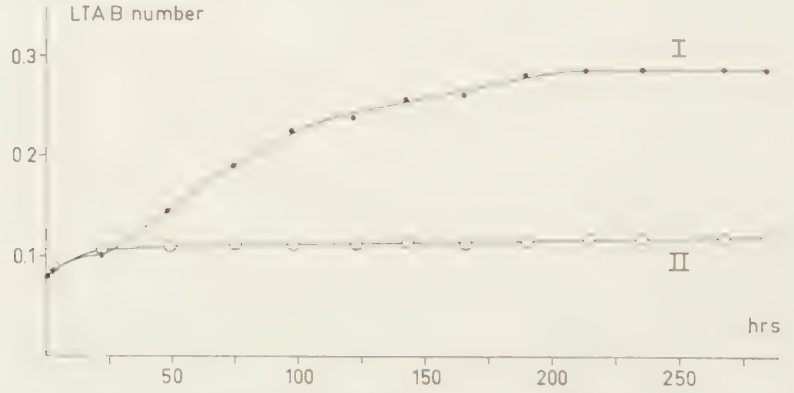

Fig. 11. Change of the LTAB number with time at pH 5,64 in two series of closed tubes. At the start of series I contains the normal concentration, series II a much reduced concentration of dissolved oxygen.

In series I, where originally the normal concentration of dissolved oxygen was present, the surplus negative charge increases with time so that a value of the LTAB number is reached of about 0.29. Compare curve I.

[1]) For preparation of this 2 % sol see beginning of section 11.

[2]) See Lunge-Berl, Chemisch-technische Untersuchungsmethoden, Julius Springer, Berlin 1921; Erster Band p. 568.

[3]) The LTAB number in these experiments is found by dividing a_{opt} by four, because we have used 2 ml 2 % sol instead of 2 ml 1 % sol. Compare Section 2, paragraph e.

In series II, having originally about 1/10 of the oxygen concentration of series I, a nearly horizontal level is reached much sooner. This horizontal level is situated at a much lower LTAB number (about 0.13). Compare curve II.

These results strongly support the explanation given above for the horizontal levels found in the case of closed tubes as represented by figs. 9 and 10. That the increase of the surplus negative charge up to its final level is accompanied by a consumption of oxygen has been proved by determining the dissolved oxygen. We found for the tubes of curve I and II that at the end of the experiment practically no oxygen was left in solution. This, in our opinion, demonstrates that the increase of the surplus negative charge with time is due to oxydation by oxygen.

11) Influence of the tertiary butanol concentration in the sol mixtures

Starting from the same solution of purified egg-phosphatides in 40 % tertiary butanol as was used in section 8, a primary 1 % phosphatide sol containing 10 % tertiary butanol was made by mixing 1 volume with 3 volumes H_2O. With the aid of the blowing device described in part I of this series [1], the volume was reduced by half, after which we have a 2 % sol which is free from tertiary butanol. By diluting 1 volume with 1 volume H_2O we obtain thus a 1 % sol free from tertiary butanol.

We now made in closed 8 ml tubes 5 series of sol–LTAB–H_2O–acetate buffer–30 vol. % tertiary butanol mixtures, in which only the tertiary butanol–end–concentration varied (0–3–6–9–12–vol. % respectively). After 145 hours the stoppers of the closed tubes were removed. The results are given in fig. 12.

It is seen that with all five series we reach the same horizontal level (the dissolved oxygen being used up here) in the closed tubes. The renewed increase with time of the surplus negative charge, following the removal of the stoppers (indicated by an arrow in fig. 12), gives no curves which coincide. Taking into account that there is no regular succession in these curves and some cross one another, we think it very probable that the lack of coincidence is only the result of larger errors in the estimation of the optimal flocculation (the number of flocculated tubes here being large already). It seems, therefore, that tertiary butanol has no influence on the rate of oxydation producing extra surplus negative charge.

In contradiction with this conclusion seems to be the influence of tertiary butanol in the first 20 hours. Here the series containing 12 % tertiary butanol shows a very rapid initial increase with time, whereas the series containing 0 % tertiary butanol shows no increase with time in the first hours. As in fig. 12 the succession of the curves in the first hours of the experiment cannot come clearly to the fore, we give this information in fig. 13, a graph in which the scale of the abscissa is taken much larger. We observe in this figure, that in the presence of 0 % and

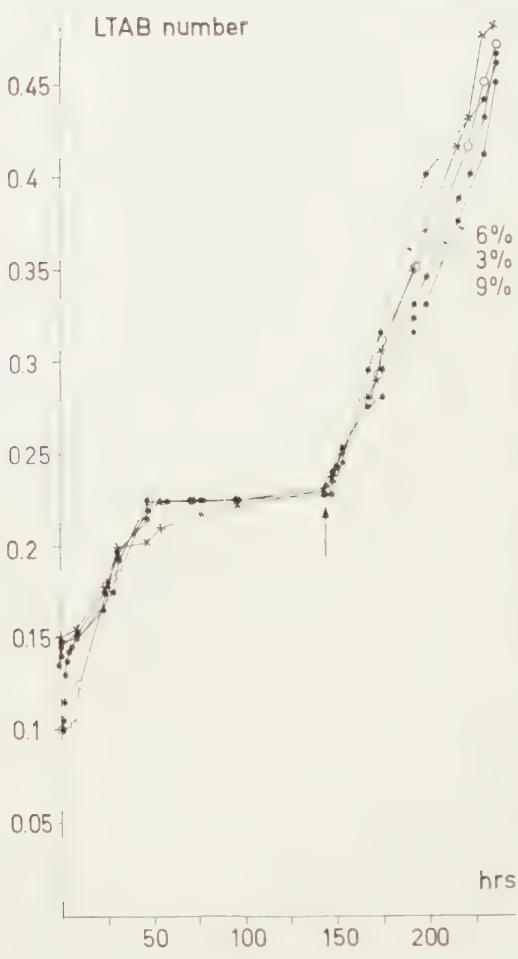

Fig. 12

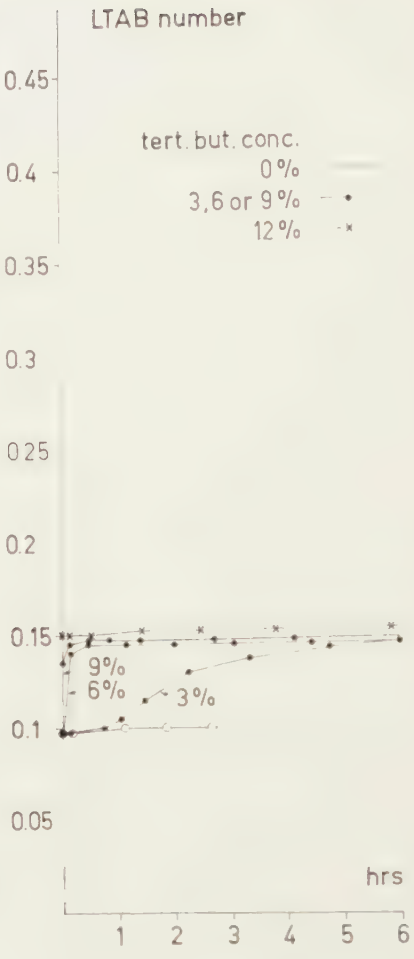

Fig. 13

Fig. 12. Change of the LTAB number with time at pH 4,65 in five series of closed tubes; the tertiary butanol concentration of these flocculation series being 0–3–6–9 and 12 vol.% respectively. At the arrow the stoppers have been removed. Explanation see Text.

Fig. 13. Change of the LTAB number of the five flocculation series of fig. 11 during the first 6 hours. The common level of the LTAB number without contact with air reached ultimately by all five series is given by a horizontal broken line. Explanation see Text.

3 % and 6 % tertiary butanol the initial LTAB number is about 0.10. This value remains for about 6 hours the same in the case of 0 % tertiary butanol; it begins to increase after 50 min. in the case of 3 % tertiary butanol. After 8 minutes it has already strongly increased in the case of 6 % tertiary butanol. The first observation of the series with 9 % tertiary butanol already gave a LTAB number of 0.135 which after 30 minutes reached 0.145 which value further remained constant. The first

observation of the series with 12 % tertiary butanol gave 0.15 which did not change any further in six hours.

It seems unlikely that the above must be ascribed to a strong promotion of the production of extra surplus negative charge by oxydation with increase of the tertiary butanol concentration, for we do not find this back in the further course of the curves (see above the discussion of fig. 12). We rather think that the influence of tertiary butanol, coming to the fore in fig. 13, is not due to a strong increase of the oxydation rate but to a physical change of the sol particles.

We propose the explanation following below, which starts from the assumptions:

a) The sol particles at 0 % tertiary butanol carry not all surplus negative charge at their surface, but part of it (here one third) in their interior.

b) The sol particles at 0 % tertiary butanol have a consistency which does not allow LTA-cations to diffuse rapidly into their interior, nor allows the substance(s) carrying the surplus negative charge to move to the surface.

c) Tertiary butanol leads to a change of the above mentioned consistency (swelling) such that the very small permeability is increased with increasing concentration.

In fig. 13 we meet two levels, the lower at a LTAB number of about 0.10 and the higher at about 0.15. The first named level then represents the surplus negative charge present at the surface of the original sol particles.

The higher situated level represents all surplus b negative charge, thus of the part originally present at the surface plus the part originally occluded in the interior and not accessible to the LTA-cations.

The increased permeability as a result of increasing concentrations of tertiary butanol then results in the relative position of the curves in fig. 13, in which the switching over from the lower level to the higher level can be seen distinctly at 3 % and at 6 % tertiary butanol.

With higher concentrations the switching over occurs so rapid that we may only observe its last stage (with 9 %) or it may escape observation (with 12 %).

The LTAB number at 0 % tertiary butanol remains constant more than six hours, it thereafter begins to increase rapidly, passes the level at 0.15 and reaches the same higher situated level as at 3, 6, 9 and 12 % tertiary butanol (compare horizontal broken line in figs. 12 and 13). Probably oxydation products lead here to increased permeability, so that the originally occluded surplus negative charge can find LTA-cations also.

If the above explanation is true, it follows that for determining the total surplus negative charge a small tertiary butanol concentration should be present.

12) Influence of ascorbic acid and of cysteine

The influence of these additions has been investigated in open tubes and in closed tubes.

In the case of open tubes we first made a series 20 ml sol mixtures (formula see section 8) and then pipetted these into a series of tubes, previously provided with 0.1 ml of a buffered ascorbic acid (or cystein) solution, 10 ml of each mixture. Thus we obtained two series of open tubes, one without and one with the above mentioned addition.

In the case of closed tubes, we first made 20 ml sol mixtures in open tubes. With each mixture we filled two 8 ml tubes to the brim, one of which was previously provided with 0.1 ml of the buffered ascorbic acid (or cysteine) solution. We then closed the tubes and thus obtained two series of closed tubes, one without and one with the abovementioned addition.

ascorbic acid

As the pK of ascorbic acid is 4.17, we must dissolve ascorbic acid and $NaHCO_3$ in water in the molecular proportions 4:3 to obtain an ascorbic acid solution of the same pH (4.65) as the sol mixtures buffered with the acetate buffer, so that we may add 0.1 ml to the sol mixtures without changing the pH. We thus made a solution of 0.440 gr. ascorbic acid 0.1575 gr. $NaHCO_3$ in 25 ml H_2O, of which 0.1 ml was added for the series at pH 4.65 with ascorbic acid. Control measurements of the pH indeed showed that the mixtures with and without ascorbic acid within the experimental errors (about 0.02 pH) showed the same pH.

The results have been given in fig. 14. We observe that in the series

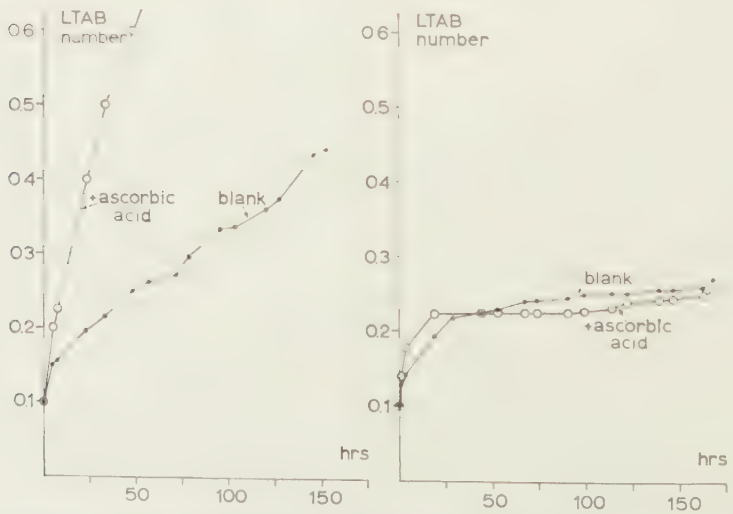

Fig. 14. Influence of ascorbic acid at pH 4.65 on the LTAB number and its change with time. Left graph in open tubes, righthand graph in closed tubes. Explanation see Text.

with open tubes the rate of the increase of the surplus negative charge with time is strongly increased by ascorbic acid. Compare lefthand graph in fig. 14. In the series with closed tubes, within the experimental errors, both curves reach the same nearly horizontal level. Compare righthand graph in fig. 14.

CYSTEINE

We started here from cysteine hydrochloride. Its solution, of which 0.1 ml was added to the sol mixtures, was made of 0.1576 gr. cysteine hydrochloride + 0.084 gr. $NaHCO_3$ dissolved in 25 ml H_2O. Control pH measurements showed that after adding 0.1 ml of this solution to the sol mixture buffered with acetate did not change the pH within the experimental errors.

The results have been given in fig. 15. We observed that neither in the series with open tubes (left hand graph) nor in the series with closed jubes (right hand graph) cysteine has an influence exceeding the experimental errors.

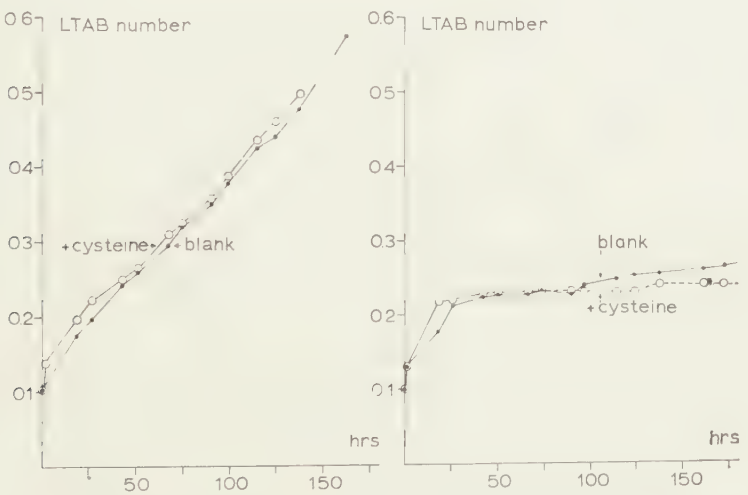

Fig. 15. Influence of cysteine at pH 4,65 on the LTAB number and its change with time. Lefthand graph in open tubes, righthand graph in closed tubes. Explanation see Text.

Discussion

The results show that cysteine and ascorbic acid, both being "reducing" substances, do not prevent phosphatides from oxydation by air (in the open tubes proceeding continuous and in the closed tubes until the dissolved oxygen has been practically used up). It is most interesting that at the pH investigated ascorbic acid in the series of open tubes even greatly increases the rate of increase of the surplus negative charge with time, whereas cysteine has no influence. Ascorbic acid thus acts as a redox-catalyst, but cysteine does not. This must mean that the redox

potential of the system ascorbic acid — dehydroascorbic acid is higher than that of "phosphatide" and its (first ?) oxydation product. The inactivity of cysteine may either mean that the redox potential of the system cysteine cystine lies lower than that of "phosphatide" and its oxydation product, or that the system cysteine–cystine cannot act reversibly with sufficient velocity.

We have also investigated whether ascorbic acid will also increase the rate of the increase of the surplus negative charge with time at pH 6.85. Within the experimental error no influence could be found. It therefore depends on the pH whether ascorbic acid acts as redox catalyst or is indifferent.

13) Deterioration of pure lecithin and cephalin in air

a) *Methods.*

Lecithin and cephalin (phosphatidyl ethanolamine) are the main phosphatide components in egg phosphatides. The question may be asked whether the increase of the surplus negative charge of our sol in contact with air is due to the formation of acidic substance(s) from one or both of these phosphatides. We have attacked this question with the aid of chromatography, starting from fresh column fractions of pure lecithin and pure cephalin obtained by one of us (H). Both fractions have been brought to a concentration of 0,3 %.

Chromatography has been performed in a small slit-feeding apparatus [2] with as mobile phase di-isobutylketone — acetic acid — H_2O (50:25:5), and using strips of 22 × 4,5 cm of washed $SiCl_4$-treated Schleicher and Schüll 2043b Paper [3], the strips having been cut in such a way that the mobile phase ascends in the direction of the smaller suction rate. Each strip has two starting points on a horizontal line 2 cm. above the immersion line i. The front was always allowed to rise 18 cm above the immersion line. After drying the chromatogram with warm air, it was stained with the improved mixed staining bath described elsewhere [4]. This bath contains the ingredients for tricomplex-staining and also Brilliant Green in appropriate concentration.

Amphoionic phosphatides, as for instance lecithin and cephalin, are stained red, but substances of acidic nature stain green. On the dried chromatogram stained spots or areas are circumlined with a pencil. Thereafter the chromatogram is treated 2 minutes with a 0,05 % $KMnO_4$ solution and then washed out with tap-water and dried with warm air.

After this treatment all red and green colour of the circumlined spots and areas disappear to make place for a brown colour.

The green background colouration of the paper also disappears and makes place for a much weaker light brown colour.

One may now observe on this slightly brown background area, which are stained brown more strongly. The contours of these areas, which were not visible after staining with the mixed staining bath, have been given on the chromatograms with broken lines.

b) *Lecithin.*

We take four strips and bring up on the two starting points of each strip 5 mm^3 of the 0,3 % lecithin solution. One strip is run directly (after drying of the solvent) with the mobile phase and thereafter stained and treated with $KMnO_4$ solution as described above. Compare strip A in fig. 16. The other three strips are suspended in the air for one, four and eight days respectively before running, staining and

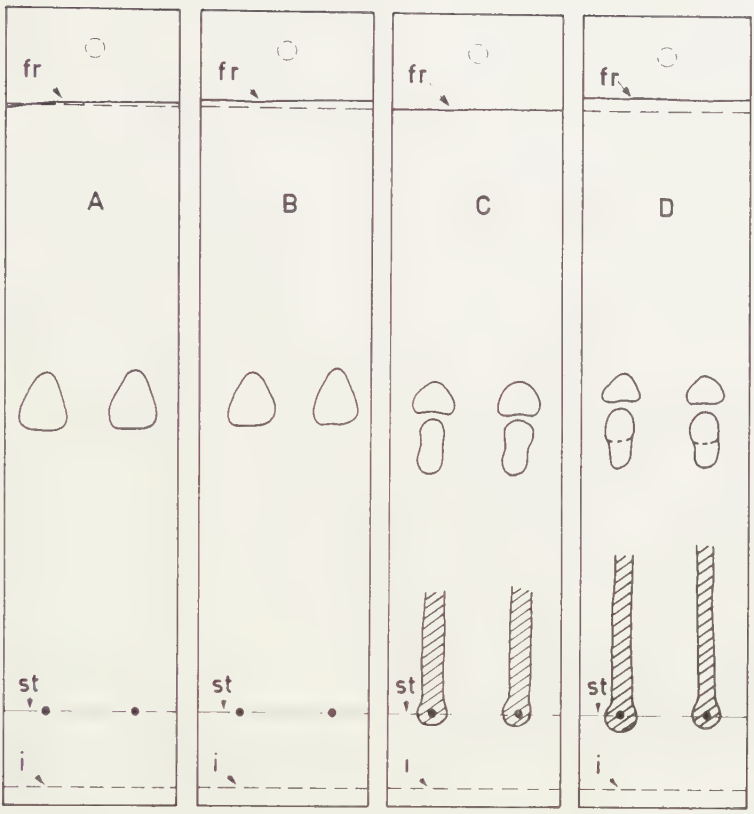

Fig. 16. Chromatograms stained with the mixed staining bath, showing the deterioration of pure lecithin in contact with air. Red spots are circumlined, green spots and streaks are shaded. Explanation see Text.

subsequent $KMnO_4$ treatment. Compare strips B, C and D in fig. 16. On the first, fourth and eight day we have also run blank chromatograms, that is, directly after applying the lecithin solution. The appearance of these three further blanks were exactly the same as strip A in fig. 16. This proves that during 8 days the lecithin dissolved in methanol/chloroform (1:4) and kept in the refrigerator (5° C) shows no signs of deterioration.

We now proceed to discuss the changes which take place with the dry lecithin on the starting points exposed to air (fig. 16, A → B → C → D). The spots which were stained red in the mixed staining bath are circumlined only; areas which in this bath are stained green are circumlined also, but shaded. The blank (A) shows only one red triangular spot. After one day in the air (B) the staining results are practically the same. The red spot has still practically the same area. After 4 days in the air the red lecithin spot is smaller, thus part of the lecithin has now been converted. We observe a weaker red, elongated spot below the lecithin spot. Apart from this, we observe a green stained streak stretching in upward direction from the starting point, which also is stained green.

After 8 days in the air, the changes which were visible after 4 days have continued. The lecithin spot is smaller than after 4 days, the red elongated spot below it is more distinct and possibly consists of two adjacent spots. The green streak proceding upwards from the green starting point, stretches higher upwards. In neither of the

four strips did $KMnO_4$ treatment show details which were not visible already after staining with the mixed staining bath.

c) *Cephalin.*

Four strips have again been taken, on the starting points of which 5 mm^3 of the cephalin solution are applied. One strip has been run directly (strip A in fig. 17), and is, after staining, and also after circumlining with a pencil, treated with $KMnO_4$. The three others were suspended in the air for one, three and 7 days respectively before running, staining and subsequent $KMnO_4$ treatment. Compare strips B, C and D in fig. 17. If we compare fig. 17 with the analogous figure for lecithin, fig. 16, it is seen that the deterioration of cephalin in air partly resembles that of lecithin, partly takes another course.

As in the case of lecithin, the area of the unchanged cephalin spot decreases with the time of exposure in air, but already after one day the area is distinctly smaller than in the blank. We also find, after three days, a green streak stretching in upward direction from the green spot at the starting point. This phenomenon is more pronounced after seven days.

Differences with the deterioration of lecithin are:

1. No new red spot below the phosphatide spot is formed.

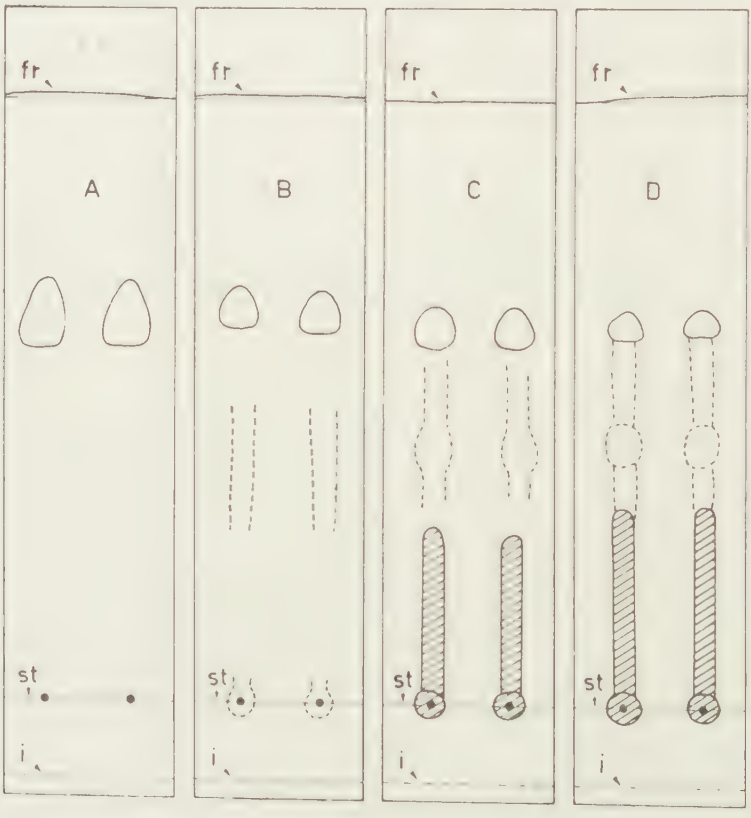

Fig. 17. Chromatograms, stained with the mixed staining bath, showing the deterioration of pure cephalin in contact with air. Red spots are circumlined, green spots and streaks are shaded. Parties which became visible after subsequent treatment with 0,05 % $KMnO_4$ are given with broken lines.

2. The treatment with $KMnO_4$ subsequent to staining reveals details below the phosphatide spot which were not visible on the stained chromatograms. (Compare the broken lines on strips B, C and D of fig. 17).

d) *Results with lecithin and cephalin compared.*

The results show that pure lecithin and pure cephalin in contact with air develop acidic substances stainable by Brilliant Green on the chromatograms. The nature of these substances remains unknown. We can only say that fatty acids are not present here, for the fatty acids move with the front. We would have observed them both with the mixed staining bath as well as with the subsequent $KMnO_4$ treatment. We have the impression that cephalin deteriorates faster than lecithin.

It is remarkable that with lecithin a red spot appears below the lecithin spot (fig. 16), whereas this does not occur in the case of cephalin. Thus among the deterioration products of lecithin substances are formed which still show tricomplex staining, which means that the phosphate-choline group is still present. One of the fatty acid of a lecithin may have lost part of the carbon chain by oxidation at the double bond(s), and thus has an OH group or an aldehyd group at the end.

Among the deterioration products analogous phosphatides with shortened fatty acid residues may be expected, having a carboxyl group at the end. They will have acidic character and may stain green in the mixed staining bath.

The fact that no new red spot forms in the deterioration of cephalin indicates that deterioration products with intact phosphate-colamine groupings do not occur. Both with lecithin and with cephalin the Schiff reaction already after one day in air is very strong. The difference mentioned above between lecithin and cephalin, suggests that certain aldehydic substances which are formed react with the NH_2 group of cephalin. The reaction product will then have an acidic character, as the phosphate group is still intact.

The above suggestions would fit in with an observation of section 8, namely that added heptylaldehyde strongly accelerates the increase of the surplus negative charge of our egg phosphatide sols in contact with air. It may also explain why our cephalin fraction dissolved in chloroform/methanol keeps reasonably well for some weeks, although the solubility of oxygen in organic liquids is much higher than in water. The alcohol present may form acetals with any aldehyde formed and may thus prevent or very strongly decrease the action of aldehydes on the NH_2 of cephalin.

14) The LTAB number of egg-phosphatides isolated from boiled hen egg. Influence of the pH

A hen egg was boiled (20 min.), the coagulated yolk (17.9 gr) was isolated and finely devided in acetone. The suspension was brought into a column. The yolk particles were then extracted with 50 ml acetone to remove the oil. Thereafter we eluted the phosphatides with 55 ml ethanol. The phosphatide solution in ethanol was evaporated, dissolved in 15 ml ether and precipitated with 40 ml acetone.

The precipitate was dissolved in 15 ml ether, this solution evaporated as a thin layer on the wall of a rotating cylindrical flask. The phosphatide (1.35 gr) was dissolved in 30 % tertiary butanol to prepare a 3 % solution, which served for the preparation of the 1 % sol. With this sol we determined the LTAB number both at pH 4.65 and at pH 6.85, and followed its change with time. Compare fig. 18. Whereas in all preceding experiments, in which we started from the commercial Merck preparation, the initial

surplus negative charge was higher at pH 6.85 than at pH 4.65, we find here practically no difference between the two.

A further point of difference is that with the commercial preparation the rate of the increase of the surplus negative charge with time is much greater at pH 6.85 than at pH 4.65. We find here practically the same rate (which is somewhat lower than that of commercial preparation at pH 4.65). We must conclude that in the commercial preparation already some deterioration is present and that it contains acid substances, which are nearly undissociated at pH 4.65 but partly dissociated at pH 6.85.

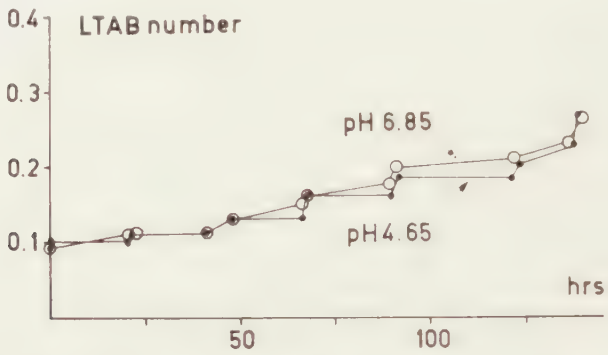

Fig. 18. Change of the LTAB number with time at pH 4,65 and pH 6,85 respectively of an egg-phosphatide sol, prepared from egg-phosphatides isolated from hen eggs. Explanation see Text.

Remains the fact that at pH 4.65 both the commercial egg phosphatide and that freshly prepared from the hen egg show about the same initial LTAB number, namely 0.1. This must be due to an acidic substance which is dissociated at this pH.

Rhodes and Lea [5] found that in egg-phosphatides a small percentage of inositides (inositol phosphatides) is present, for which they give a value of 0.6 % expressed in percentage of the total lipid phosphorus. We remind that the surplus negative charge (LTAB number) is expressed in 10^{-4} equivalent per g phosphatides. A value of 0.1 for the initial surplus negative charge thus means 10^{-5} equivalent per g.

When we take for the mean molecular weight of egg-phosphatides 750 and calculate the surplus negative charge as a percentage of total lipid phosphorus we obtain 0.75 %, which comes near to the value given by Lea and Rhodes for inositides.

15) Discussion

The results in the preceding sections lead to a two-fold origin of the surplus negative charge in phosphatide sols:

1) A surplus negative charge due to a small percentage of a strongly acid component, possible inositol phosphatides (section 14).

2) An extra surplus negative charge increasing with time, mainly as a result of oxydation.

We will not exclude the possibility that some hydrolysis also takes place. The long chain fatty acid split off does not contribute to the surplus negative charge at pH 4.65, but markedly at pH 6.85 (compare section 8). This may explain that at pH 4.65 a nearly horizontal level is reached in the closed tubes, (compare fig. 9), whereas at pH 6.85 the LTAB number continues to increase slowly with time (compare the "level" in fig. 10).

It seems that oxydation (the above point 2) sets in at the double bonds and a great variety of substances are formed (section 13).

Fatty acids with a relatively short chain have no influence on the LTAB number (section 8) but relatively short chain aldehydes certainly contribute to the surplus negative charge both at pH 4.65 and pH 6.85 (section 8). A suggestion to explain this effect has been given in section 13 (reaction with cephalin).

After splitting off a short chain aldehyde, the remaining part of the phosphatide molecule may possess a carboxylic group, (section 13), which will contribute to the surplus negative charge only markedly at pH 6.85; hardly at pH 4.65. In the commercial preparation a beginning of deterioration in the latter sense is already present (LTAB number greater at pH 6.85 as at pH 4.65). In freshly isolated egg-phosphatide practically no COOH groups are present (LTAB number equal at pH 6.85 and 4.65).

With the present communication we end the series of "Hydrosols of Total Egg-phosphatides".

Although we have given a rapid and convenient method for the preparation of Sols in Part I of this series, we have learned in the present communication which precautions should be taken (section 4) to prevent an increase of the surplus negative charge during the preparation. For further study of the colloid chemistry of phosphatide sols, it will, however, not be useful to start from a commercial preparation. One should isolate the phosphatides and perform experiments with their sols under conditions which exclude oxydation (nitrogen atmosphere).

Summary

1. The negative charge carried by egg-phosphatides in hydrosols has been investigated by determining the maximum flocculation in the flocculation-zones obtained in low disperse sols with linoleyl trimethylammonium bromide (LTAB).

2. We call the abovementioned negative charge the surplus negative charge. It has been found that LTAB is quantitatively bound to the surplus negative charge at maximum flocculation.

3. The surplus negative charge in sols of commercial egg-phosphatides and of egg-phosphatides freshly isolated from boiled hen eggs amounts

at pH 4.65 about 10^{-5} equivalents per g phosphatides. This must be due to a strong acid component, possibly inositol-phosphatides.

4. The influence of various factors on the surplus negative charge in sols of the commercial preparation has been investigated (pH, temperature, time, additions such as H_2O_2, oleic acid, heptylic acid, heptyl aldehyde, ascorbic acid, cysteine).

5. With commercial egg-phosphatide the surplus negative charge is higher at pH 6.85 than at pH 4.65, which is not the case with egg phosphatides freshly isolated from hen eggs. Associated with the former are therefore acidic substances of weakly acidic character, which must have been produced by a beginning deterioration of the amphoionic phosphatides.

6. The surplus negative charge of sols increases with time, (the flocculation zone shifts in the direction of higher LTAB concentrations). It has been proved that this is due to oxydation by oxygen dissolved in the sol. Without contact with air the shift comes to a standstill after a certain time (the dissolved oxygen has then nearly been used up), but resumes its original rate when the sol is now brought in contact with air (oxygen dissolves in the sol).

7. Paper chromatography has confirmed that pure lecithin and pure cephalin in contact with air give rise to substances with acid character. Cephalin reacts faster than lecithin.

8. The relatively strong increase of the surplus negative charge with time, in sols at pH 6.85, must be ascribed partly to oxydation products weakly acidic in character (COOH-groups). The smaller increase with time at pH 4.65 cannot be explained in this way as the weakly acidic substances are practically not dissociated at this pH.

9. In connection with the increased production of surplus negative charge in sols by added heptylaldehyde, not only at pH 6.85 but also at pH 4.65, it has been suggested that aldehydic substances formed react with cephalin giving rise to strongly acid reaction products, (the basic group of cephalin being attacked, the phosphategroup remaining intact).

10. To study the colloid chemistry of phosphatides it will be necessary to isolate the phosphatides and to perform experiments with their sols under conditions which exclude all oxydation (N_2 atmosphere).

Department of Medical Chemistry,
University of Leiden

REFERENCES

1. Bungenberg de Jong, H. G. and J. Th. Hoogeveen, these Proceedings, Series B, **62**, 201 (1959).
2. ——— and ———, these Proceedings, Series B, **63**, 1 (1960).
3. ——— and ———, these Proceedings, series B, **64**, 1 (1961); see section 2.
4. ———, these Proceedings, series B, **64**, 445 and 458 (1961).
5. Rhodes, D. N. and C. H. Lea, Biochem. J. **65**, 526 (1957).

CONVECTION-CURRENTS IN THE MANTLE OF THE EARTH

BY

F. A. VENING MEINESZ

Vanuxem lecture, held at Princeton, December 15, 1958.
Revised text

(Communicated at the meeting of May 27, 1961)

For many reasons, which we shall presently examine, we have to assume that the Earth's cooling episodically causes convection-currents in the mantle; in our period they are no doubt present. Their vertical dimensions vary from the whole thickness of the mantle, i.e., up to 2900 km depth, to smaller depths involving not more than a few hundred kilometers. In the present period their horizontal sizes are of the order of the vertical ones or slightly larger. Their velocities, at least in this period, seem to vary from a few cm/year to about 10 cm/year.

There is reason to suppose that the whole mantle for the greatest part consists of $(Fe, Mg)_2SiO_4$, for a smaller part of $MgSiO_3$ and possibly other constituents, all in crystalline state. The flow must, therefore, have the character of a more or less continuing deformation in crystals which, according to modern research, is a perfectly acceptable hypothesis. This hypothesis is in line with the fact that we must assume that practically over the whole thickness of the mantle the flow is of the plastic type, by which term we want to indicate that the differences of the principal stresses have to exceed a certain limit before flow movement can occur. According to the theory of Huber–Hencky–von Mises this criterion can be expressed by the condition that the sum of the squares of the nine stress constituents of the stress deviator [1]) must surpass a limit which we may indicate as the limit of elasticity. We can also express it by the equivalent condition that the elastic deformation energy has to exceed a limit if by elastic deformation energy we understand the energy connected with elastic change of shape exclusive of change of volume.

Although it is probably a simplification of a more complicated phenomenon, it is usually assumed that the velocity of flow, occurring if the limit of elasticity is passed, is proportional to the excess of stress above this limit, the excess for each constituent of the stress deviator being supposed

[1]) The stress deviator is derived from the stress tensor by diminishing the normal-stress constituents by their mean value and so the sum of the three constituents of the principal diagonal of the stress deviator is zero. The stress deviator is, therefore, given by five quantities, the stress tensor by six.

to occur in the same ratio to the limit of this constituent. The elastic deformation, which has little importance as soon the flow deformation far exceeds it, is usually supposed to remain proportional to the total stress.

Volume changes of an elastic kind may in general be neglected and probably also those of a more or less permanent character, but two effects on volume have great importance as far as they affect the density, the thermal expansion effect and the effect of a change of crystal state. They, in fact, provide the driving power of the convection-currents.

It is interesting to see that modern research in solid state physics has shown that up to a certain temperature the elastic limit diminishes with rising temperature but that it does not continue to do so for a further temperature increase: for magnesium, e.g., it remains constant for further increasing temperatures. This behavior checks well with the experience obtained for the earth. There is no evidence that below the rigid crust, i.e., below 35 km, the mantle shows any decrease with depth of the limit of elasticity although this limit does show a strong decrease with depth in the rigid crust. At the surface it may be estimated that the elastic limit for uniaxial stress is of the order of at least 10^9 dynes cm^2 (1000 kg cm^2) but at the lower boundary of the rigid crust, i.e., at a depth of 35 km., its value cannot be more than 10 20 . 10^6 dynes cm^2 (10 20 kg cm^2) and it seems to be like that over the whole mantle. The first of these latter statements is derived from the deviations from isostasy which are observed over areas of more than 1 Mm^2 (10^6 km^2); the mean is seldom above 10 to 20 mgal. The conclusion that this limit of elasticity does not vanish in the lower part of the mantle can be deduced from the fact that there is no evidence that lower mantle currents start easier than currents in the surface layer.

After this short introduction concerning the possibility of currents in the mantle, we shall proceed to the arguments in favor of the hypothesis that they episodically occur, that in this period they are present, and that they have the character of convection-currents caused by the earth's cooling. We shall afterwards cope with the problem how such currents can break through the layer between the depths of 500 and 900 km where the density, after Birch's reduction for the effects of pressure and temperature, shows a transition from the value of about 3.3 present above the depth of 500 km to that of about 4.0 valid for the lower 2000 km of the mantle.

Strong arguments in favor of large size convection-currents in the mantle are provided by the island-arc areas. Taking as an example the Indonesian Archipelago, we may notice that the belt of strong negative gravity anomalies, which may be interpreted as a belt of crustal down-buckling, indicates uniaxial compression in the crust in about the same direction throughout the whole archipelago; it does so by the difference in size of the anomalies in the different parts of this belt as well as by its

pattern of distribution which can only be understood if we assume uniaxial compression in a direction NNW–SSE. Now such a stress field can easily be explained by supposing a subcrustal current of great horizontal dimensions flowing out from under the Asiatic continent, but the theory of general crustal contraction cannot account for it nor the assumption of local currents subsiding below the downbuckling belt. The size of the subcrustal current is such that we must assume it to be a current system throughout the whole depth of the mantle.

Similar conclusions may be drawn from the gravity field in the area of Japan and in the region of the Caribbean. For the latter case, the subcrustal current must be supposed to flow out towards the NNE from below the South American continent.

All indications, as well from the geological as from the geophysical side, point to the anomaly belt in these areas representing geosynclines in the making. The further history of geosyncline belts provides us with still more evidence of subcrustal currents and points to their having an episodical and half turn character. As GRIGGS has first mentioned, these properties can be accounted for by the combination of two facts, firstly the limit of elasticity of the mantle matter, i.e., its plastic behavior, and secondly, the extremely slow temperature conduction in the mantle. This conduction is so slow that during the time of 50–100 million years the mantle current phenomenon lasts, the temperature conduction may be neglected. As a consequence of this we may assume that the temperature is carried along with the flowing matter and also the density corresponding to that temperature.

This leads to the following picture. Because of the earth's heat radiation towards the outside, an inward temperature gradient develops, causing dynamic instability, the upper layer being denser than the lower. However, as long as the direction of the temperature gradient coincides with gravity, this cannot lead to a current system because the elasticity limit connected with the mantle's plastic behavior prevents movement. If, however, a secondary phenomenon, e.g. in the surface layers, brings about a horizontal temperature gradient of such magnitude that the resulting horizontal pressure gradient can overcome the elasticity limit, a flow movement sets in which must take the shape of a revolving convection current of a size more or less dependent on the horizontal dimension of the pressure gradient area at the surface; if we may consider the whole phenomenon as two dimensional, the depth, though also dependent on boundary conditions, will tend to be about equal to the horizontal size.

As the current proceeds, the subsiding column will contain more and more of the surface layer of lower temperature and, consequently, higher density, and the rising column more and more higher temperature matter of lower density. The maximum density difference between subsiding and rising columns will come about when the current has made a quarter revolution and so the velocity of the current must then be largest. From

then on, the driving power of the current decreases again, and its velocity, therefore, also; the current comes to a stop after having made a half turn. The low temperature matter is then at the bottom and the high temperature matter on top; dynamic stability is re-established. It may be remarked that the elastic limit, perhaps, causes the movement to stop slightly before it has made an entire half turn; this, however, does not amount to much in the general picture. We may conclude that the mantle convection-currents have a half turn character.

After dynamic stability has been restored, the temperature conduction combined with smaller convection systems, must gradually change the situation; the uppermost layer by radiation towards the outside, loses its higher temperature and the bottom layer of lower temperature is heated by the core, which thus loses much more heat than would otherwise be possible through conduction if the mantle were stable. It may be noted here in passing that this provides us a new argument in favor of mantle convection currents: the inward temperature gradient needed for bringing about the currents in the core, now assumed to explain geomagnetism, is otherwise difficult to account for.

It follows from our hypothesis that at the beginning this stage of the mantle's history must be characterized by high temperature gradients at the earth's surface. It continues until the original situation is restored; as we shall discuss, this process must be accelerated by smaller convection systems. The upper layers of the mantle have then again lower temperature than the lower layers and the whole phenomenon can repeat itself. We thus find the explanation of the episodic character of the half turn convection currents in the mantle and of the orogenic periods of the crust caused by them. We thus obtain another important argument in favor of the mantle-current hypothesis. The fact that these currents take place over the whole thickness of the mantle proves that the matter in the lower part of the mantle is plastic and has an appreciable limit of elasticity.

From the above considerations, it is clear that the phenomenon cannot be exactly periodic; the secondary surface phenomenon needed to put the convection current in motion, will, in general, not recur after equal periods. Assuming the maximum velocity of the current to be of the order of 10 cm/year, we find the total duration of a current system over the whole mantle to be of the order of 50 to 100 million years and this checks with the geological findings concerning periods of orogeny. The value for the maximum velocity of 10 cm year has been derived from the average relative velocity of adjoining crustal blocks separated by fault zones and shockwise moving during earthquakes, as e.g. the movement along the San Andreas fault in California.

The picture of half turn convection as it has been developed gives a good explanation of the regression of the continents occurring during the first stage of an orogenic period. Because of the greater radioactive content of sialic matter as compared to the crustal matter below the oceans, we

may assume the temperature in the mantle below the continents to be higher than below the oceans and this may well give rise to the horizontal temperature gradient needed for starting the half turn convection current. There must exist, therefore, a tendency toward a mantle current pattern with the currents rising below the continents and flowing off towards their borders; the current bringing about the crustal deformation in the Indonesian Archipelago, as well as those causing the other island arcs along the east coast of Asia, all correspond to this pattern. As, however, we saw that the rising current contains mantle matter of higher temperature than the subsiding current we can understand that during the flow of the current a thermal expansion takes place below the continents, especially below their border areas, and a thermal contraction below the adjacent parts of the oceans. Obviously, this must bring about a regression of the continents. After the currents have made a quarter turn, this regression must gradually diminish, and disappear when the currents come to a stop. This later development, therefore, accounts for the transgression occurring in the period the tectonic activity dies out. Here again the mantle current hypothesis gives a satisfactory explanation.

We must assume, that also below some areas in the oceans rising mantle-currents occur. As the writer exposed elsewhere [1]) this may be supposed to be the case below mid-ocean ridges. The same vertical movements of the crust as mentioned above, probably occur in the belts of these ridges, which during an orogenic period thus come into being and afterwards disappear again.

The stopping of the mantle-current must cause a decrease or a disappearing of the uniaxial compression in the crust and this must lead to a tendency towards readjustment of the isostatic equilibrium in the downbuckled belts mentioned above for some island-arc areas. In the continents the geosyncline belt, where the crust was thickened by the compression, may rise to a high folded mountain range. In the oceanic areas, however, the crust has nearly the same density as the mantle on which it floats, and so the readjustment of the isostatic equilibrium can only lead to the rise of a low submarine ridge.

Continuing now the history of the continental geosyncline we may realize that the folded mountain range must in two ways disappear. At the surface it is attacked by erosion, but at the lower boundary of the crust, the downbuckled root must also gradually disappear. The temperature conduction must slowly raise the temperature of this root to the value of the layer it has been pressed down in and render it plastic. It will then start to flow out under the foreland and lift up the crust to an increasing distance from the original geosyncline belt. Applying this to the Alps, it is obvious we thus can explain the coming into being of the French and German "Mittelgebirge".

[1]) e.g. in a paper "Continental and Ocean-floor Topography; Mantle Convection-currents, Revised Paper", Proc. Kon. Ned. Akad. v. Wetensch., Ser. B (1961).

This hypothesis solves also another problem. The folding and overthrusting in the Alps have led to an estimated crustal shortening of 200 to 300 km. In connection with the thickness of the rigid crust of 35 km and assuming a crustal shortening of 250 km, this must lead to an excess of mass of a cross-section at right angles to the belt of 35×250 equals 8750 km^2. Now the cross-section of the Alps may be estimated roughly at about 150 km in breadth and an average of 2.5 km in height, i.e. at 375 km^2. If we assume that the undeformed crust is constituted of two layers of equal thickness, one consisting of granite and one of basalt, the cross-section of the root of the Alps is 6.52 times larger than that of the Alps themselves and so the total excess of matter in each cross-section is 7.52 − 375 equals 2820 km^2. This is so much less than the figure we derived from the compression over 250 km, that it is difficult to account for by erosion only. If, however, this Alpine root flowed off below the crust and lifted it up, thus forming the European Mittelgebirge, we may add the whole amount of mass-excess represented by these Mittelgebirge and their roots, which, moreover, must also comprise a great part of the matter eroded from the Alps. We thus arrive at a figure for the mass-excess coming very near to that given by the 250 km compression. The small surplus may easily have been carried away by erosion and river transport. These considerations give a strong support to our hypothesis, but we have to explain why, at least for the western part of the Alps, the Mittelgebirge only rose up in France and Germany, i.e. on the outside of the Alpine arc, and not also on the Italian side, i.e. on the inside of it. We have also to make clear how the subcrustal matter could attain to such great distances from the Alps. Both points can be simply explained if we assume a mantle-current system flowing out towards the western and northern edges of the European continent and carrying along the Alpine root matter.

Such a current-system is also made likely by the geological evidence, e.g. in the area between the North Sea and the Alps. In the southern part of the Netherlands the surface of the Tertiary and lower layers show "horsts" and "grabens" in a SE–NW direction which can easily be accounted for by assuming crustal tension in the direction at right angles to it; the tilted fault planes delimiting the horsts are diverging downwards and those delimiting the graben are converging in that direction. The relative vertical movements can, therefore, be explained as a simple isostatic equilibrium readjustment. Similar features are found in the whole area between the Netherlands and the Alps and we may see larger features of the same kind in the Upper Rhine Graben between the Black Forest and the Vosges and in the Lower Rhine Graben in the neighborhood of Bonn. The crustal tension in SW–NE direction in this part of Europe can evidently be explained by the current-system mentioned above, which diverges from the Alps towards the French, Belgian, Dutch and German coasts.

The fact that these "horst" and "graben" formations – other well-known examples of graben development are of course found in the rift belt of Africa and in the Dead Sea area in Palestine – originate in the same period during which, elsewhere, horizontal compression causes folding and overthrusting, obviously is strong evidence for the existence of mantle current systems and against the contraction hypothesis. We may add that the continuing shockwise movements along fault planes, as e.g. the San Andreas and other faults in California and the Great Glen fault in Scotland, are pointing in the same sense. All this evidence makes it probable that the earth crust consists of large blocks, separated by zones of weakness and affected by the drag forces exerted on them by mantle-currents.

Still another argument in favor of great convection systems in the mantle can be derived from Bowen's theory about the formation of basalt. Bowen assumes that pressure relief brings about selective fusion of peridotite, giving basaltic magma and olivine; the latter remaining in the solid state. However, the selective fusion of peridotite in the upper layer of the mantle must soon have caused the layer to lose its basaltic constituents and become barren. The problem then arises how the continuation of the formation of basalt throughout the whole history of the earth can be accounted for. It is obvious that episodically recurring convection currents in the mantle can provide an explanation, each current having brought new peridotite to the surface, producing new basaltic magma. It is even possible that the slight increase of density caused by the change of peridotite into olivine has contributed to the energy needed for the flow of the convection currents.

A strong argument in favor of mantle-currents is provided by the evidence given by geomagnetism which points to great shifts of the earth's rotation axis with regard to the crust. According to Runcorn, the North Pole, since Carboniferous times, must have described a curve from the middle of the Pacific via Japan towards its present site. This movement cannot have been a shift of the rotation axis in space; with regard to space, it must have remained stable but it must have meant a movement of the earth's crust relative to the earth's interior. If we assume systems of convection currents in the mantle, such a relative movement can easily be explained; it is unlikely that the drag exerted by such a current system on the crust would have no resulting moment causing such a relative movement.

It is true that if the matter of the mantle were a Newtonian fluid, the drag exerted on the crust by a convection-current system caused by the earth's cooling would indeed have no moment around an axis through the earth's center, but the fact, that the stress deviator working in the mantle must surpass an elastic limit before flow can originate, makes for asymmetries in the current-system which practically must lead to the presence of a drag-moment.

As an instance of such an asymmetry it may be remarked that in many cases the convection-current must have a one-sided character. In case e.g. the crustal compression leads to a geosyncline in which the sialic crust is thickened, the concentration of sialic matter, which is richer in radioactive constituents than more basic rocks, leads to a higher temperature in the upper mantle. This brings about a rising current below the geosynclinal belt and we may assume that this usually will only lead to a convection cell on one side of the belt. If towards that side the elastic limit is overcome and a cell forms, the limit towards the other side will never be reached and no second cell on the other side can originate. If the geosynclinal belt has a curved pattern the cell will doubtless come into being on the inside of the curve because such a cell meets somewhat less resistance than a cell on the convex side of the curve; the latter will therefore not form.

This brings us to the discussion of a number of facts pointing to smaller convection currents below the crust in island-arc areas. In the Indonesian Archipelago as well as in the West Indies we find deep basins inside the island-arcs of recent origin, and it is likely that also in other island-arc areas, the deep basins have not long ago been formed; Kuenen and Fallot have been able to prove that the Mediterranean area south of the French and Italian Riviera have in a late period subsided below sea-level while at the same time the Riviera area has been subject to strong rising. These relatively quick changes of level are probably to be attributed to convection systems in a layer reaching down from the crust to a depth of between 400 and 1000 km. Considering e.g. the Banda Sea area in the Indonesian Archipelago, we know by the evidence of sedimentation products on the surrounding arc, that some 1 to 2 million years ago the basin area must have been above sea-level. It now shows depths over 5000 meters, while in the central part submarine ridges are present. A convection system in this area can be easily explained in the way mentioned above; the surrounding tectonic arc is characterized by strong folding and overthrusting in the upper Miocene and so we may here expect a strong increase of crustal matter of higher radioactive content than the adjacent crust shows. The resulting higher temperature must lead to a temperature gradient from the basin towards the arc which, in connection with the general vertical cooling gradient, must cause a convection cell subsiding below the basin and rising below the arc. The convection is one-sided; no second cell develops on the outside of the arc. In the way already dealt with for the subject of regression and transgression, this can indeed explain a subsidence of the basin area, but the fact that the depths are so large for all the deep basins in the Indonesian Archipelago is interesting; it must probably be attributed to the effect on this process of the transition layer in the mantle. We shall come back to this point.

The one-sided character of the convection, however, is not limited to

the smaller cells causing the deep basins in island-arc areas. It is no doubt likewise true for cells of continental dimensions reaching down through the whole mantle. In view of the island-arcs along the southeast and east coast of Asia we may probably conclude that the great convection below that continent mainly flows out toward these coasts. The South American continent seems to be dragged towards the NNE with regard to the Caribbean area. The relative movement in California of the North American continent with respect to the North Pacific crust appears to have a strong component to the SSE i.e. along the San Andreas and other more or less parallel faults.

We shall only shortly mention here a last argument in favor of convection currents in the mantle. We shall afterwards come back to it in another paper in this same number of the proceedings and show that a correlation exists between the shapes and distribution of the continents and convection-currents in the mantle. This makes it likely that such currents have played a decisive part in the originating of continents and oceans. However, if they did so during the first part of the earth's history, during which period the continents must, at least for the greater part, have formed, there seems little reason to assume that convection currents no longer occurred in more recent periods; the earth's cooling, which is their principle source of energy, still goes on.

We shall herewith conclude our argumentation in favor of convection-currents in the mantle. The writer thinks the evidence is rather conclusive. We shall now take up the problem of how these currents can break through the layer of density transition between 500 and 900 km depth and we shall find that this is not only possible, but that the presence of the transition layer even favors them. It increases their driving power to more than four-fold the value it would have if they only were brought about by an inward temperature gradient.

The thermodynamic and other problems connected with the transition layer have been dealt with by E. J. W. Verwey, Director of the "Philips Physical Laboratories" at Eindhoven, Netherlands, and J. L. Meijering and C. J. M. Rooymans, connected with these Laboratories. The two latter scientists have written an important study [1]) on these problems which comes to the conclusion summarized in the following.

A treatment of the mantle transition layer as a transition of a one component system, $(Mg, Fe)_2SiO_4$, from the orthorhombic to the spinel phase does not yield satisfactory results. Clapeyron's law, given by

$$\frac{dT}{dP} = \frac{dV}{dS}$$

in which T is the absolute temperature, P the pressure, V the volume and S the entropy, provides a temperature gradient downwards of 3.6°/km

[1]) Meijering, J. L., and C. J. M. Rooymans, "On the Olivine-Spinel Transition in the Earth's Mantle"; Proc. Kon. Ned. Akad. v. Wetensch., Ser. B, **61**, 5, 1958.

over the whole transition layer and this is too much; besides other considerations, the fact, that the density discontinuity in the core at a depth of about 5100 km is probably caused by a change of the iron to a solid state, fixes the temperature at that depth at about 4000°, which is incompatible with the above temperature gradient.

A treatment of the matter in the transition layer as a binary system, Mg_2SiO_4–Fe_2SiO_4, gives us better results, but it is not satisfactory either, as it makes it probable that in this case the phase transition would split up in two fairly thin layers near the upper and lower boundaries of the transition layer, and this does not concur with the seismic data about this layer.

The authors come to the conclusion that the matter in the transition layer must at least be a ternary system, probably Mg_2SiO_4, Fe_2SiO_4, $MgSiO_3$, but that it is likely that still more constituents are present, viz. Na, K, Ca, Al etc. The constitution of the mantle would then consist of the following four layers, of which the two middle ones would together form the density transition layer:

1. *Olivine + Enstatite*, in which the crystal lattice would not dissolve the above named other constituents (Na, K, Ca, Al, etc.). It would, therefore, be a heterogenous layer and its physical properties would be an average of those of its constituents.
2. *Spinel + Olivine + Enstatite* ⎫ transition layer; the density curve
3. *Spinel + Enstatite* ⎭ may show one or more kinks.
4. *Spinel* in the whole layer of the mantle below the transition-layer.

The spinel lattice is very tolerant and so NA, K, Ca, Al, etc. can occupy cation vacancies in this lattice.

It is clear that this state of things must lead to reactions if the pressure changes while the temperature remains the same. For a loading of the crust, part of the olivine will change into the denser crystal phase, i.e. into spinel, and this must be accomplished by the release of a transition-heat, which we shall denote as K cal gram. Evidently this transition must obey Clapeyron's law which says

$$\frac{dP}{dT} = \frac{KA}{T\Delta V}$$

where P is the pressure, T the absolute temperature, ΔV the change in volume and A the heat-equivalent equalling $4.19 \cdot 10^7$. We shall not here attempt a solution of these thermodynamical problems, but we may restrict ourselves to the remark that an unloading of the crust, e.g. by the disappearing of an iceload, must thus be accompanied by three phenomena at the crust's surface, viz. a rapid elastic rebound, a somewhat slower rising because of the reaction by the transition layer (change of spinel into olivine) and a still slower rising caused by the plastic flow of the mantle matter toward the unloaded area.

The evidence about the postglacial rising of Fennoscandia seems to show that the first phenomenon was nearly instantaneous; that the main part of the second phenomenon has taken place in 2 to 3 thousand years, while the last phenomenon, during the lapse of 10,000 years since the unloading, has caused about two-thirds of the deviation from isostatic equilibrium to vanish.

In this fairly simple case of flow in the mantle the observed facts allow a surmise about the part played by the transition-layer. For convection-currents in the mantle the problem is much more complicated. The authors, already mentioned above, in their important paper on the constitution and the thermodynamic problems of the mantle, state that in their opinion convection-currents in the mantle can break through the transition-layer and that the presence of this layer even favors their coming into being. In a study of the writer of this present paper, he came to the conclusion that probably this layer causes the driving power of such a convection-current to be about 4 times increased, compared to what would happen if the mantle were homogeneous and if the convection-current, therefore, would only derive its driving power from a downward temperature gradient.

There is still another phenomenon in which the transition-layer probably plays a part, viz. in the fairly rapid subsidence of deep basins inside island-arcs. We have already discussed these basins and attributed them to smaller convection cells rising under the arc and subsiding under the basin. We drew attention to the great depth of these basins considering the small size of the current. It is clear that the presence of the transition layer may provide an explanation. The subsiding current must have brought about the change of a considerable amount of olivine into the spinel phase, and this must have caused a stronger subsidence than the difference in temperature between the rising and subsiding columns in normal convection could have brought about.

CONTINENTAL AND OCEAN-FLOOR TOPOGRAPHY MANTLE CONVECTION-CURRENTS REVISED PAPER

BY

F. A. VENING MEINESZ

(Communicated at the meeting of May 27, 1961)

The spherical harmonic development of the Earth's topography can provide us with important information about the geophysical history of the Earth. In 1922 PREY [1]) published such a development up to the 16th order. Little was done with the results. In 1951 the writer undertook a study of these results and found that they can lead to valuable conclusions if we combine the coefficients of the $2n + 1$ underterms of the term of the nth order in one representative figure, representing the root of the mean square of the elevations represented by all these underterms of the nth order. We thus obtain a curve of 16 figures, which are invariant with regard to a change of the coordinate system.

The important views obtained led to the undertaking of a new development, pursued up to the 31st order, in which the greatly increased knowledge about the ocean-floor topography could be taken into account. Made possible by a subsidy of the Netherlands Organization for Pure Scientific Research, it was carried out, under the supervision of Dr. Ir. D. J. HOFSOMMER and Mr. M. L. POTTERS, by the Mathematical Center in Amsterdam. Prof. Ir. G. J. BRUINS of the Geodetic Laboratory of the Polytecnical University at Delft with his collaborators took care of the map reading, needed for the 40680 figures of the topographic elevation required for the development. For obtaining the representative figure t for each order n the root was taken of the mean square of the elevations represented by the $2n + 1$ underterms of this order, for which the formula (11–50) was used mentioned on page 426 of "The Earth and its Gravity Field" by W. A. HEISKANEN and the writer (McGraw–Hill).

The upper curve of fig. 1 shows the representative figures t thus obtained for the orders 1 to 31. The middle curve gives similar representative figures s for a second development, carried out for the ocean-floor topography, putting zero all the figures of positive elevation, i.e. the land topography figures. The lower curve shows the representative figures l

[1]) PREY, A.: Darstellung der Höhen- und Tiefenverhältnisse der Erde durch eine Entwicklung nach Kugelfunktionen bis zur 16. Ordnung, Abhandl. Ges. d. Wiss. Göttingen, Math.–Phys. Kl., N.F., vol. 11, no. 1, 1922.

for a third development, made quite recently of the land topography, putting zero all the figures of negative elevation, i.e. the ocean-floor topography. The coefficients of this last development were not separately determined but they were simply found by taking the differences of the coefficients of the first and the second development. As a consequence of this, the representative figures l are slightly less accurate than the

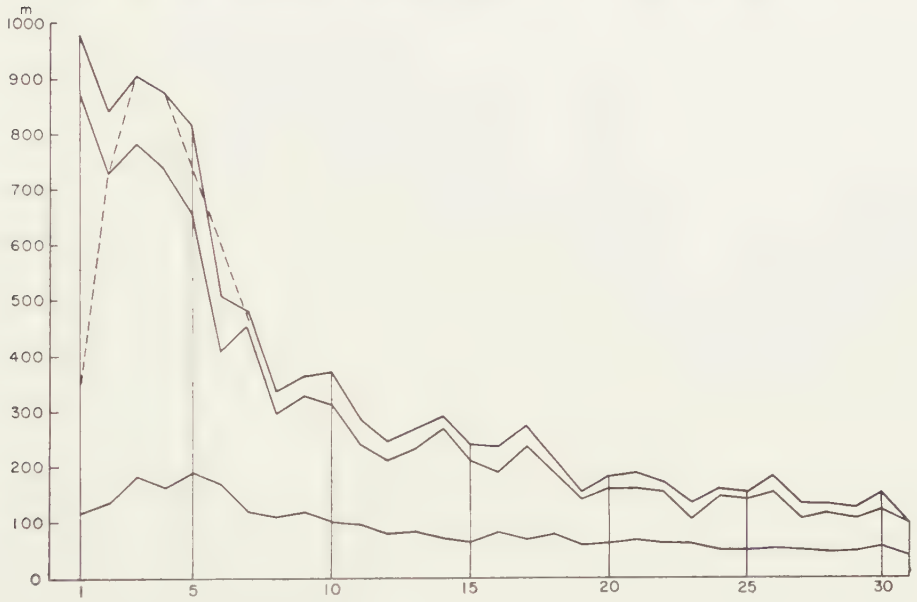

Fig. 1. Spherical harmonic development of the Earth's Topography; representative figures. Upper curve: total topography t; middle curve: Ocean-floor topography s; lower curve: Continental topography l.

values of t and s; the mathematical center gave as an estimate that errors of 10 are possible. The development thus carried out for the land topography led to much smaller values of l than those formerly derived from t and s; the writer suspected these latter values of l and this was the reason for requesting the mathematical center to derive l directly from the coefficients of the spherical harmonic development. This led to this revised paper, in which the writer shall repeat statements of former papers in order to supersede these papers as completely as possible.

Figures 2 shows three curves, T, S and L, which are derived from those of figure 1 by multiplying the ordinates by $n^{\frac{1}{2}}(n+1)^{\frac{1}{2}}$. In this way the ordinates for higher values of n become larger and their study is made easier. A multiplication by n would have served the same purpose; the above factor was chosen for hydrodynamic reasons, treated shortly on page 428 of "The Earth and its Gravity Field".

From the curves of fig. 1, as well as from those of fig. 2 we can in two directions draw conclusions. We can study the common characteristics of the curves and we can examine the differences between the ocean-floor

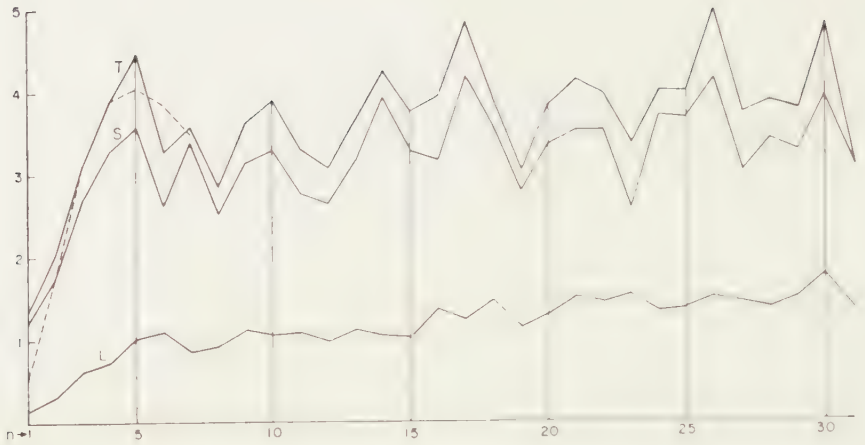

Fig. 2. Ordinates of Fig. 1, multiplied by $n^{\frac{1}{2}}(n+1)^{\frac{1}{2}}$, giving T, S and L.

curves and the continental curves. On behalf of the more easy comparison, we have in fig. 3 given a curve of the ratio $q = l/s = L/S$.

The following table gives the quantities t, s, l, T, S, L and q.

Table of t, s, l, T, S, L, q (t, s, l in meters)

n	t	s	l	T	S	L	q	n	t	s	l	T	S	L	q
1	979	869	116	1385	1229	164	0.133	16	239	192	84	3942	3166	1386	0.437
2	841	729	135	2060	1786	331	0.185	17	275	239	71	4810	4181	1242	0.297
3	905	783	182	3135	2712	631	0.232	18	212	192	80	3921	3551	1480	0.417
4	875	738	164	3913	3301	734	0.222	19	156	143	59	3041	2788	1150	0.413
5	815	657	190	4464	3599	1041	0.289	20	186	163	63	3812	3340	1291	0.387
6	509	409	170	3299	2651	1102	0.416	21	192	163	70	4127	3504	1505	0.429
7	481	454	119	3600	3398	890	0.262	22	175	156	64	3937	3509	1440	0.410
8	338	300	111	2868	2546	942	0.370	23	143	110	65	3360	2584	1528	0.591
9	384	332	120	3643	3150	1138	0.361	24	163	150	54	3993	3674	1323	0.360
10	373	316	103	3912	3314	1080	0.326	25	156	143	53	3977	3646	1352	0.371
11	289	243	96	3320	2792	1103	0.395	26	186	156	56	4928	4133	1483	0.359
12	247	212	80	3085	2648	999	0.377	27	135	110	52	3712	3024	1430	0.473
13	271	235	84	3656	3170	1133	0.358	28	135	119	48	3847	3391	1368	0.403
14	293	271	73	4246	3926	1058	0.269	29	127	110	50	3746	3245	1475	0.455
15	243	212	66	3765	3284	1023	0.311	30	156	127	57	4757	3873	1739	0.449
								31	101	101	42	3181	3181	1322	0.416

We can in the first place remark the curiously regular character of the curves, which do not at all make an at random impression. The values of neighbouring ordinates do not jump up and down. It is clear that this regularity can only be caused by some physical processes behind the origin of the Earth's topography.

One of the most important parts of this topography is found in the transition zones between the continents and the oceans, which usually

occupy only narrow belts. Most if not all of this topography is comprised in the ocean-floor topography indicated by s and at most only a small part in the continental topography l.

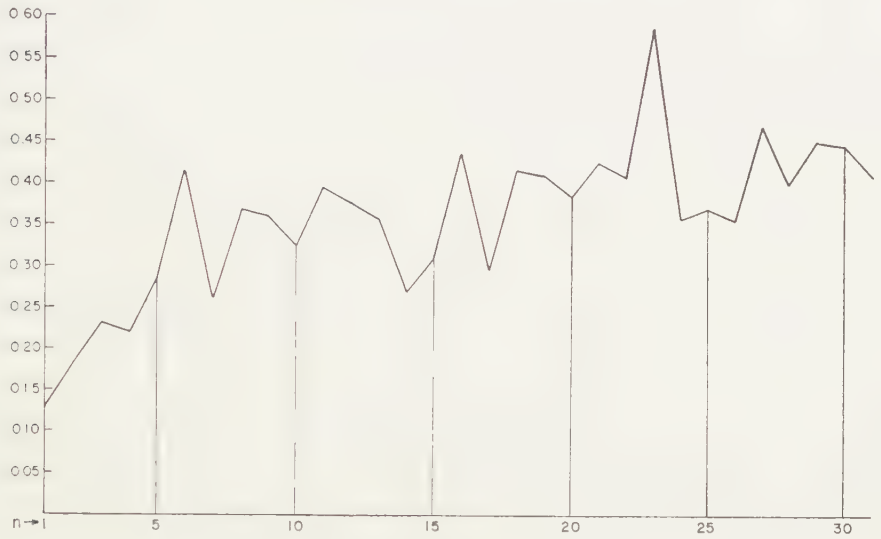

Fig. 3. Ratio $q = l/s = L/S$.

Obviously other topographic features of the ocean-floor are incorporated in s and of the continental topography in l. About the continental topography we may remark that it is strongly affected by erosion and by sedimentation. In the oceans erosion does not take place and sedimentation mostly near the coasts though turbidity currents (KUENEN and his collaborators) may bring the sediments to great distances.

Examining fig. 1 we see that in all curves the ordinates for values of n larger than 5 diminish with increasing n. For the continental curve the ordinates are much smaller than for the others, but they decrease slightly less for increasing n. As increasing values of n must correspond to diminishing horizontal dimensions, we can no doubt explain the decrease of the ordinates as a proof that in general smaller topographic features are less high than larger. That for the continental topography this is somewhat less the case than for the oceanic topography is probably caused by the fact that in the continents the erosion tends to increase the smaller topography while in the oceans the sedimentation tends to diminish it. For features of continental dimensions – which correspond to values of n up to five – the behaviour is different; in general the ordinates do not show a decrease.

The large first order terms in the curves for t and s are evidently due to the asymmetric distribution of the continents. This checks with the fact that these curves practically entirely comprise the great topographic feature represented by the transition belts between oceans and continents; in the curve for l the first order term is not especially conspicuous. In

§ 11. 5 of "The Earth and its Gravity Field" the writer gave the following hypothesis for accounting for the asymmetric distribution of the continents.

According to this hypothesis it is attributed to a convection-current during a very early stage of the Earth's history, when it was still an undifferentiated fluid body of high viscosity. As fig. 4 shows, this current,

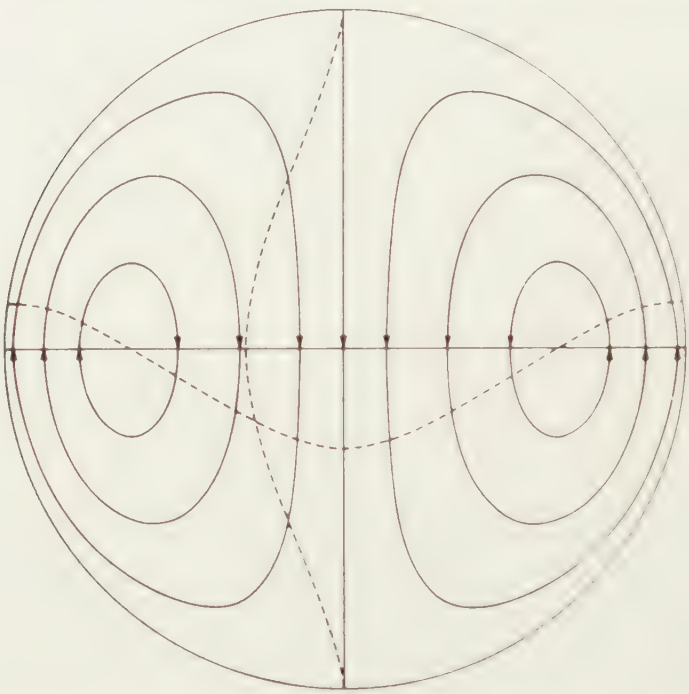

Fig. 4. First-order current in a fluid undifferentiated earth during early stage of its history (figure taken from „the Earth and its Gravity Field", p. 423); dashed curves give velocity distribution in axis and equator of current-system.

caused by the cooling, must have passed through the center of the Earth and returned along the surface. In this way it brought the heavy matter to the core and the light sialic matter to the surface, where the current concentrated it in an urcontinent above the subsiding current. This caused the asymmetry of the continental distribution. By the process described, the core, the mantle and part of a sialic crust came into being.

The urcontinent must have solidified rather quickly, because in doing so, though becoming denser, it remained lighter than the mantle matter and kept floating. The mantle surface, on the contrary, must for a much longer time have remained fluid; if at the surface parts solidified, they became denser and subsided and thus were melted again.

As the cooling continued, new convection-currents must have originated in the mantle, and we may expect these currents to have drawn the rigid urcontinent apart and to have transported the parts over the mantle

surface till they came to a stop above areas of subsidence of the currents. A group of independent continents thus came into being, showing similarities of shape on both sides of the rift oceans dividing them.

We may expect that at some time of the Earth's history the mantle surface between the continents also solidified; further relative movements of the continents must then have become more difficult. It is not easy to form an opinion at what time this occurred. As we shall deal with afterwards, the relative movements of continents still continued and probably even go on in the present period, but the velocity of movement must have been diminished.

Returning to our interpretation of the curves of t and s in figures 1 and 2, we shall see that these curves give a remarkable confirmation of the hypothetic geophysical history here set down. Examining these curves, we see that they show a series of waves, which for higher values of n are best shown by fig. 2, but for smaller values up to $n=8$ by fig. 1. Neglecting at first the peaks for $n=7$, to which we shall presently come back, we see that these waves show their maxima at $n=3$ or 4, 9 or 10, 14, 17, 21, 26 and 30. Before dealing with all these waves, we shall start by treating of the first, which covers the range from $n=2$ till $n=6$, or if we neglect the peak for $n=7$, till $n=8$. With this exception the curves of t and s in fig. 1 are regular. They show a clear correlation to the dashed curve of which the ordinates are inversely proportional to the Rayleigh numbers for mantle-currents in a mantle consisting of a Newtonian viscous fluid, of which the vertical velocity components are distributed according to a spherical harmonic of the same order as that of the ordinate. For the deriving of these Rayleigh numbers the writer may refer to § 11–4 of "The Earth and its Gravity Field", where it is shown that these numbers only depend on the order n of the distribution and not on the $2n+1$ subterms of which the nth order term consists. With this dashed curve the writer intended to represent the probability that a mantle-current is distributed according to these orders. The scale of the ordinates of the curve was chosen in such a way that the ordinate for $n=3$ coincided with the curve for t.

The correlation of the dashed curve of fig. 1 to the curves for t and s is a strong argument in favour of the hypothesis that the urcontinent was drawn apart by mantle-currents and that such currents determined where in the course of the Earth's history the continents were located. In this connection we may stress again that the s and t curves practically contain the whole coastal topography between continental blocks and sea-floor areas.

There can not be much doubt, that in the present period currents are going on in the mantle. The presence of earthquakes points to the crust being subject to forces, and the crustal movements occurring during earthquakes prove that these forces can not be caused by the contraction of the earth, but that they must be due to drag exerted by mantle currents.

For the following reasons we must assume that the vertical velocity components of these currents are no longer mainly distributed according to third and fourth order spherical harmonics but that the fifth and sixth order terms play the largest part.

This may be derived from the spherical harmonic development of the continental elevations which shows relatively large values for $n=5$ and $n=6$. In fig. 3 we see that the ratio $q=l/s=L/S$ for these two values of n is considerably larger than for $n=3$ and for $n=4$. This points to the effect of geosynclinal topography. As the writer has on many occasions already put forward, the horizontal compression in geosynclines of a light continental crust must after the readjustment of the isostatic equilibrium lead to a much higher topographic elevation than the horizontal compression in geosynclines of a rigid oceanic crust which in its mean density differs only slightly from that of the mantle. This leads to the well known absence of high mountain ranges in the oceans; this absence may, however, not be interpreted as a proof that geosynclines do not occur in the oceans; they may be quite numerous there.

As geosynclinal topography in the continents must have a fairly recent character because otherwise erosion must have destroyed it, we may conclude that the mantle-currents, which in the earth's crust by their drag bring about this horizontal compression, must in our present period have two characteristics, viz. a distribution according to a fifth order spherical harmonic and, secondly, a distribution according to a spherical harmonic of a sectorial character which causes crustal compression in one direction and not in a sense at right angles to it. The equator of this sectorial system does not need to coincide with the geographical equator-plane of the earth.

If we draw a cross-section of the Earth in the equator-plane of the sectorial system, which has been schematically represented by fig. 5, we see that the mantle cross-section comprises slightly more than ten circles; we find in fact that the number is 10.8. If we divide the mantle cross-section in exactly ten sectors, each sector represents the cross-section of a mantle-current, of which the vertical velocity component is distributed according to a sectorial spherical harmonic system of the fifth order. If we divide the mantle in twelve sectors, each sector represents the cross-section of a mantle-current, of which the vertical velocity component is distributed according to a sectorial spherical harmonic system of the sixth order; in this case circles tangent to each other and to the division lines of the mantle cross-section would be nearly tangent to the outer and inner boundaries of the mantle but they would remain a little inside these circles.

We can now come to an interpretation of these facts. This is based on the hypothesis that in the present period of the Earth's history the mantle has become crystalline. We require this hypothesis to explain that convection-currents reaching over the full height of the mantle, which

for many reasons, exposed by the writer in another paper in this same number of the proceedings, in chapter 11 of "The Earth and its Gravity Field" and in many other papers, we have to accept, can break through the density transition layer between 500 and 900 km depth. In the upper mantle layer we must assume an orthorhombic phase, in the lower two thousand kilometers of the mantle a cubic (spinel) phase [1]).

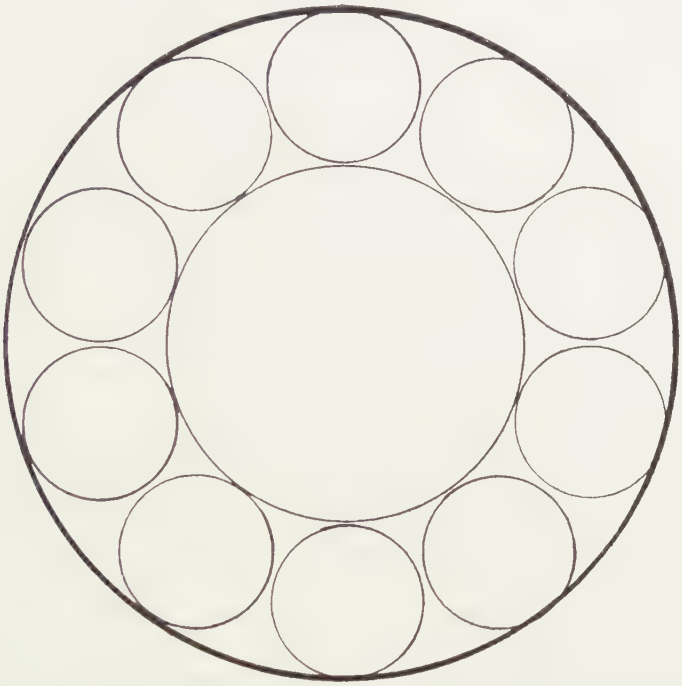

Fig. 5. Schematic representation of mantle currents in a crystalline mantle.

From the above facts it seems to follow that even in a crystalline mantle convection-currents can occur, but that their distributions differs from that which comes into being in a mantle consisting of a Newtonian viscous fluid; it appears to assume a distribution which allows the central part of a current-cell to rotate without much permanent deformation. This leads to a cross-section perpendicular to the axis of rotation of which the boundary lines are tangent to a circle. According to what has been mentioned above, this gives a distribution of the vertical velocity component according to a fifth and sixth order sectorial spherical harmonic.

We thus find that in the course of the Earth's history the mantle-currents, no doubt because of the earth's cooling and the resulting crystallization of the mantle, have changed from a viscous or pseudo-viscous type to a flow in crystalline matter of which the vertical velocity

[1]) For a more thorough treatment of these problems the writer may refer to: J. L. Meijering and C. J. M. Rooymans: On the olivine-spinel transition in the Earth's mantle, Proc. Kon. Ned. Akad. v. Wetensch. Ser. **B 61**, 5, 333–344 (1958).

component has a sectorial spherical harmonic distribution. The drag effect on the crust has, therefore, also changed its character. When the flow was still viscous or pseudo-viscous, crustal compression above areas of current subsidence must have brought about shields. In the present period crustal compression must have the character of uniaxial compression, causing a geosyncline. In case the mantle-currents bring about crustal stress release, it has likewise uniaxial character and, as it has been exposed in chapter 10D of "The Earth and its Gravity Field", we may expect a graben to develop.

The crystalline state of the mantle must also bring along an elastic limit below which only elastic deformation is possible. Only if the stress deviator exceeds this limit, permanent deformation is possible, which may get the character of pseudoflow. It is possible that during the later part of the viscous state of the mantle, its behaviour has also shown an elastic limit, which had to be exceeded before pseudoflow came about. If that is true, it is difficult to obtain an indication at what time this elastic limit originated; it may have been very early.

The presence in the mantle of an elastic limit must have given and must still give a half-turn character to the mantle-currents, bringing down the matter of low temperature caused by the cooling at the surface and bringing on top the matter of high temperature of the lowest mantle layer. The motoric force of the current is produced by the density differences caused by the temperature differences. After a half turn the dynamic equilibrium in the mantle is restored. In the present period we may estimate this half turn to last from 50 to 100 million years. A long time must elapse before the cooling at the surface has again lowered the temperature at the surface and heated up the temperature at the bottom of the mantle — the latter process is obviously caused by the hot core — until the original conditions are again present and a new half-turn current can come about. It is evident that for the phenomenon here described another condition has also to be fulfilled, viz. a temperature conduction in the mantle that is slow enough that in a hundred million years the motoric force of the current is not too much destroyed by it. Another requirement is also needed; as the elastic limit stabilizes the mantle, which, by the cooling at the surface, is rendered unstable, it is necessary that a secondary process, as e.g. a temperature gradient in a horizontal sense, occurs for setting the half turn current into motion. Once in movement, the density differences caused by the temperature differences, which both may be considered to be carried along by the current, bring about an acceleration during the first quarter revolution and a slowing down during the second quarter. Because of the temperature conduction we must assume that the half turn revolution is not quite fully realized; shortly before, the current must come to a stop.

The above phenomenon gives a good explanation of the episodic character of the mantle-currents and the orogeny caused by them in the

crust. As we are not sure that an elastic limit has already been present in the beginning of the Earth's history, we do not know whether this episodic character of the currents and the orogeny has always been present. That the phenomenon is not exactly periodic is caused by the secondary processes required for setting the mantle-currents into motion; they may differ in character and time.

The half turn mantle-current supposition as here advocated has already in 1939 been proposed by GRIGGS. In a remarkable paper [1]), based on extensive experiments [2]) [3]), he shows that in the mantle, which he likewise assumes to be crystalline, a type of half turn convection flow is possible —and probably occurring— which he denotes as pseudo-viscous. This conclusion is founded on the results of his experiments on pseudo-viscous flow in Solenhofen Limestone under strong shear-stresses and under a confining pressure of 10.000 atmospheres, which corresponds to a depth of about 35 km, i.e. of the surface of the mantle. Other experiments with alabaster under small shear-stresses and in the presence of its own saturated solution gave similar results, which have been attributed to solution and recrystallization. It followed that the velocity v of the pseudo-viscous flow, in both cases of entirely different character, was related to the stress τ according to the equations

$$\ln v = k\tau, \text{ or } v = e^{k\tau}, \tag{1}$$

in which k is a kind of viscosity constant.

Probably the first of the two cases is directly applicable to our problem of pseudo-viscous mantle-currents, but it seemed worth while to mention also Griggs's second result. The conclusion following from the equations (1) is, that for low values of shear-stresses —or in general of deviatoric stresses— under high confining pressure the viscosity is very high, but that for high deviatoric stresses the viscosity is small, which favors the coming into being of flow. Griggs shows how this property, in the same way as the elastic limit that has to be overcome before flow can start, brings about half turn mantle currents as described above.

These considerations bring our story for explaining the first wave of our spherical harmonic ordinates of fig. 1 to a close. Before discussing the peak of the s and t curves for $n=7$, we shall first undertake the explanation of the other waves with peaks for $n=9$ or 10, 14, 17, 21, 26 and 30. It is likely that we can account for them by currents in more than one layer in the mantle, each current rotating in opposite sense to the one above it and below it; together the current layers fill up the whole mantle. The peaks in the curves for s and t would consecutively

---

[1]) GRIGGS, DAVID; A theory of mountain-building, Amer. J. o. Science, **237**, 611–650 (1939).

[2]) GRIGGS, DAVID; Deformation of Rocks under High Confining Pressures, J. Geol., **44**, 541–577 (1936).

[3]) GRIGGS, DAVID; Creep of Rocks, J. Geol., **47**, 225–251 (1939).

correspond to two, three, four, five, six and seven current layers. We shall assume that all currents in the same column above each other are conformal. The lowest rows of the following table give the thicknesses d of each current layer for these six cases; the case of only one layer is also added. The mantle thickness has been put at $2900-35=2865$ km; 35 km was taken for the thickness of the rigid crust. The second row gives the ratio $180°/2n$ which is half of the angle which at the earth's center we must expect the convection-cell to subtend; we shall denote this ratio by ψ. The third row contains the values of the ratio $\frac{1}{2}d$ 6336 $\frac{1}{2}d$ in which d is here the height of the upper convection-cell and 6336 the radius of the mantle surface at the depth of 35 km, i.e. of the lower crustal boundary; this ratio equals the sine of the angle q which is half of the angle subtended at the earth's center by a convection-cell in the cross-section of which a circle can be inscribed. We shall henceforward call such cells "circular" convection-cells. The fourth row gives the angles q themselves, derived from these sines. The fifth row contains the ratio of the values ψ of the second row, divided by the values q of the fourth row. If the circular convection-cells corresponding to the layers in which the mantle is divided, would exactly fit the peaks of our spherical harmonic development of the earth's topography as given by figures 1 and 2, we must expect these ratios to equal the unity. We see that they slightly deviate from this value; we shall presently discuss these deviations.

TABLE II

n	5	10	14	17	21	26	30
$\frac{180°}{2n}=\psi$	18°.00	9°.00	6°.43	5°.29	4°.284	3°.461	3°.000
$\frac{\frac{1}{2}d}{6336-\frac{1}{2}d}$ (= sin φ)	1432/4904	823/5513	576/5760	442/5894	359/5977	302/6034	261/6075
φ	16°.97	8°.585	5°.740	4°.306	3°.444	2°.874	2°.462
ψ/φ	1.06	1.05	1.12	1.23	1.24	1.20	1.22
d	2865	1646 1219 2865	1152 942 771 2865	885 761 655 564 2865	718 638 565 500 444 2865	605 547 494 448 405 366 2865	522 479 439 404 370 339 312 2865

The result that all the current-types, derived by our hypothesis of more than one current-layer in the mantle, are nearly circular, is favourable for this supposition, but we have to realize that it gives reason for some

astonishment. Already long ago the writer assumed the presence in the mantle of small convection-cells which might explain the relatively quick subsidence to great depth of small basins in island-arcs, as e.g. the Banda basin in Indonesia, the Caribbean basin in the West Indies, the basin between Corsica and the Riviera in the Mediterranean, etc. However, he then supposed that the lower horizontal part of the current was distributed over the whole lower part of the mantle, because thus its velocity would be small and the inner friction too. Now the results of the spherical harmonic development of the topography seem to show that over the whole thickness of the mantle a column of small cells comes into being. It is likely that this is due to the crystalline state of the mantle and that these small cells rotate, more or less like cog-wheels, with their greatest part remaining undeformed. We may here remark that during the later stage of an orogenic period, when the half-turn current through the entire mantle has for the greatest part been completed, there must be two mantle layers where the temperature causes instability; the upper layer, where the cooling at the surface brings about a downward temperature gradient and the lowest layer where the half-turn current has brought matter of relatively low temperature which by the core is heated up; there also we may expect a downward temperature gradient. It is possible, though not at all sure, that both are needed to bring forward these columns of small convection-cells throughout the whole height of the mantle.

If this last supposition is right, it would lead to the interesting conclusion that the deep basins in island-arc areas would especially come into being during the later stage of an orogenic period.

The problem how we can understand the originating of these columns of small current-cells throughout the whole height of the mantle can perhaps find a solution by means of the results, mentioned above, obtained by Griggs. From the equations (1) it follows that for low values of shear-stresses or, more generally expressed, for low values of deviatoric stresses, accompanied by high confining pressures, the viscosity is very high, but for high deviatoric stresses under these conditions the viscosity is low. This result renders impossible, that, in case of a small mantle current cell at the surface, the lower horizontal current distributes itself over the whole lower part of the mantle; such a current would be accompanied by only small deviatoric stresses and, therefore, by very high values of the viscosity, which would raise the inner friction to an extremely high value. The solution, however, that the current takes place by many circular current-cells above each other, as described, gives undeformed inner parts of each cell, rotating over a half turn, and between and around this areas under strong deviatoric stress, where, therefore, the viscosity would be low and the inner friction likewise. Under these conditions we can understand, that the last solution absorbs less energy than the former, and, therefore, comes into being.

We thus see, that the results for the behaviour of crystalline matter

found by GRIGGS, lead to phenomena, which at first sight appear to be unlikely. They are no doubt very different from what occurs in viscous behaviour of Newtonian fluids.

We shall now discuss the small deviations from circular convection-cells, shown by the deviation from the unity of the ratio $p:q$ given in table II. For the value of this ratio for $n = 5$ we saw that the small deviation may well be explained by the presence of a sixth order term, which in the continents appears to exist. A similar explanation can not be given for the small deviation of the ratio for $n = 10$, nor for the larger deviations shown by the ratio for the values of $n = 14$, 17, 21, 26 and 30; for the latter four cases the mean of the deviations from the unity is 0.22. This means that these cells are decidedly broader than high.

It seems possible that for these cases the transition layer in the mantle between 500 and 900 km plays a part. We see that for these cases the center of the upper layer of cells, which are found at depths of 442, 359, 302 and 261 km, is situated above the transition layer and so we may conclude that the lower horizontal parts of these currents are entirely, or at least partly, located in this transition layer. For the third value of the ratio this is only partly true and for the two first values it is not the case at all. The question, therefore, is whether the coincidence mentioned can account for the broadening of the cells.

Probably this is the case. Below the subsiding column of a convection-cell the pressure is higher than below the rising column; this pressure difference explains, in fact, the horizontal flow from the first to the latter. During this flow a decrease of pressure must, therefore, occur and the result must be a corresponding phase-change from the spinel to the orthorhombic phase. The resulting density decrease must involve an increase of volume, and as the adjacent cells below the upper cell must oppose the expansion downwards, we may assume that this must mainly occur in a horizontal direction, i.e. in the direction of flow. This effect may well explain the horizontal broadening of the cell. This possibly accounts for the four last values of the ratio $p:q$ which are somewhat larger than the unity.

The curious conclusion, dealt with here, of the coming into being of columns of as many as seven cells above each other, which is difficult to account for without these cells having the character of rotating cells with central parts rotating as solid crystalline matter, seems to give some new views on the behaviour of crystalline matter subjected to strong stress. The peaks in the curves for t and s, which led to these conclusions, are only partly shown by the curve for l: probably this is due to the fact that the erosion relatively quickly affects the continental topography, especially for smaller features.

We still have to examine the peaks in the t and s curves for $n = 7$. As for this value of n the continental curve l shows a small value, we may

probably conclude to an oceanic feature. For two reasons we can not consider it in the same light as the peaks just mentioned which also are lacking in the curve for *l*. In the first place this peak can not be brought into connection with a mantle-current; it does not show in the dashed curve of fig. 1 for a current over the whole mantle height. In the second place the dimensions of the corresponding topography appear too large for making it likely that in the continents it can be taken away by erosion.

These large dimensions make it probable that this peak refers to the topography of mid-ocean ridges, as e.g. the Mid-Atlantic ridge and the Mid Indian Ocean ridge. A stronger indication in this sense is provided by the fact that this peak is not present in the spherical harmonic development of the earth's topography by PREY; at the time this development was published these mid-ocean ridges were not yet discovered. At that time only the islands on these ridges were known. The enormous increase of our knowledge of the ocean-floors, provided by the echo-sounding method, has revealed them.

On the two mid-ocean ridges mentioned Ewing and his collaborators have discovered graben and this points to tension in the crust. The same conclusion follows from the great number of volcanoes on these ridges, partly visible as islands, partly submarine volcanoes. In connection with the fact that we may suppose the Atlantic and the Indian Oceans to have originated when, in an early stage of the Earth's history, the urcontinent was drawn apart by mantle-currents, the supposition may be advanced that these ridges somehow are still due to rising currents in the mantle, which cause tension in the crust above them. A confirmation of this hypothesis is provided by the strong heat-flow found above the crests of these mid-ocean ridges; we may expect that rising currents have higher temperature than normal.

A further argument in favour of our supposition is found in the explanation it can give of the distribution of these ridges according to a seventh order spherical harmonic. If topography comes into being above the subsiding columns of a mantle-current system as well as above the rising columns, we may expect the distribution of this topography to obey a spherical harmonic of twice the order as that which governs the vertical velocity components of the current itself. As in the beginning of the Earth's history the orders 3 and 4 were the most likely for the current distribution, we can understand that the mid-ocean ridges are distributed according to a 7th order spherical harmonic. A further study would, nevertheless, be necessary to make sure about this conclusion.

An explanation that in an early period of the Earth's history ridges originated above rising mantle-currents may probably be given by the supposition that in this period the mantle-currents still transported light sialic matter towards the Earth's surface. Most of this light matter was no doubt added to the concentration of such matter above the subsiding limbs of the mantle-currents, thus forming the continental shields.

A smaller part of the sial above the central columns of the rising currents may, however, have remained there, being more or less in equilibrium between the horizontal current-limbs in the surface layer of the mantle, diverging towards both sides. These parts must thus have formed the cores of ridges, floating on the mantle in the areas between the continental floes, i.e. the cores of mid-ocean ridges.

These sialic cores account for the fairly low seismic velocities found on the ridges' crests, which usually range from 4.0 to 5.5 km sec. However, the hypothesis, here sketched, can not provide the complete explanation. In long and much appreciated discussions with my friend H. H. Hess of Princeton University he gave me several arguments for the view that the mid-ocean ridges are ephemeral features, probably present during an orogenic period. One of the strongest arguments is the existence, derived by Hess from his great many soundings of atolls and guyots, of a mesozoic ridge extending from the Mariannes arc towards Chile. It has largely disappeared since middle cretaceous time, leaving a belt of atolls and guyots which have subsided 1 to 2 km [1]). Also Menard [2]) suggests that the mid-ocean ridges may be ephemeral features.

It is not difficult to reconcile both viewpoints. Realizing that during the half-turn mantle-currents the rising column contains the mantle matter of the lower layer, which has high temperature, and the subsiding column the matter of the upper mantle layer, which has low temperature, we see that the thermal expansion of the rising column must cause a greater volume of this column and that the thermal shrinking of the subsiding column must bring about a smaller volume of this column. The result is a rising of the mantle surface — and of the crust floating on it — above the rising column and the reverse above the subsiding column. We may safely assume that the pressure gradient, which in the upper mantle layer thus comes into being, from the rising column towards the subsiding column, causes the limit of elasticity in this layer to be overcome, which leads to flow from the first column to the second. In the lower half of the mantle the reverse must occur. Below the rising column the lower boundary of the mantle must be depressed and below the subsiding column it must be elevated; the resulting pressure gradient in the lower mantle layer must be directed from the subsiding column to the rising one and must cause flow in that direction. The sketch of the total half-turn convection-current is thus completed. From our reasoning it follows that the rising of the crust above the rising mantle column must be about half of the linear thermal expansion of this column in vertical sense. Putting the coefficient of linear thermal expansion of the

[1]) Hess, H. H.; Evolution of Ocean Basins, Manuscript prepared for "The Sea, Ideas and Observations", edited by M. N. Hill, E. Goldberg, W. Munk and C. O'D. Iselin; Interscience Publishers.

[2]) Menard, H. W.; Development of median elevations in the ocean basins: Bull. Geol. Soc. o. Amer., **67**, 1623–1640 (1958).

mantle matter at 7×10^{-6} we find that the surface rising is 10 m per degree and that, therefore, a mean excess of 100° to 200° over the whole mantle column is needed for causing a ridge at the surface of 1 to 2 km as mentioned by Hess. For the highest parts of the Mid Atlantic ridge these figures have perhaps to be increased to 300° to 400°. None of these figures appears to be too high to be accepted.

Moreover, for two reasons it seems possible that a smaller excess of temperature may be sufficient to cause the mid-ocean ridges. In the first place the limit of elasticity and the pseudo-viscosity of the lower mantle layer may be less than those of the upper layer; this would reduce the depression of the lower boundary of the rising mantle column. In the second place the reduced pressure in this column might lead to a transition in the layer between the depths of 500 and 900 km of the cubic modification to the orthorhombic modification, i.e. of the spinel into the olivine modification. This would account for a further rise at the surface.

The explanation here given for the mid-ocean ridges implies that they are only present as high ridges as long as mantle currents are going on. As there can not be any doubt that in the present period this is true, this explanation can be accepted; the frequency of earthquakes and the mobility of the Earth's crust as proved by repeated geodetic measurements, are, with many geological and geomorphological arguments, a proof that we are living in a period of mantle currents.

In connection with the way the mid-ocean ridges are distributed, we may remark that in view of the low value of q (fig. 3) for $n = 14$, which is nearly as low as the value of q for $n = 7$, it might be possible that part of the peak for $n = 14$ of the curve for s of fig. 1 is also caused by these ridges. A further indication in this direction is that Prey's development of 1922 does not show such a peak for $n = 14$.

The fact that in the present period the mantle-currents are probably mainly distributed according to a fifth order spherical harmonic, and not according to third and fourth order terms, presents us with the problem why still the 7th order and the 14th order terms appear to play the dominating parts in the distribution of the mid-ocean ridges. It seems possible that this is caused by the sialic core of these ridges, which, by its larger content of radio-active matter, brings about a higher temperature in the mantle below these belts and thus favours the localization there of the rising mantle-currents, which cause the ridges to become higher.

ACTION OF CLOSTRIDIUM WELCHII TOXIN ON PHOSPHATIDES OF RED CELL MEMBRANES[1])

BY

L. L. M. VAN DEENEN, J. DE GIER AND G. H. DE HAAS

(Communicated by Prof. J. F. ARENS at the meeting of April 29, 1961)

MACFARLANE and KNIGHT (1941) discovered that the lethal hemolytic factor of Cl. welchii α-toxin (phosphatidase D) converted lecithins into phosphorylcholine and diglycerides. The rate of hydrolysis of sphingomyelins appeared to be significantly lower than that of lecithins, whereas isolated cephalins were not attacked at all (ZAMECKNICK *et al.* 1947); MACFARLANE 1948). In agreement with these observations we have recently found that several synthetic phosphatidyl ethanolamines in an aqueous medium resisted phosphatidase D action (VAN DEENEN *et al.* 1961).

On account of this information on the degradation of the individual classes of phosphatides it seems likely that the hemolytic action of the Cl. welchii α-toxin is brought about by an enzymatic hydrolysis of the choline containing phosphatides, and in particular of the lecithins present in red cell membranes. Since in red cell ghosts of the animal series sheep, ox, pig, man, rabbit and rat the lecithin content was found to increase gradually from about 0 % to 60 % of total phosphatides (DE GIER *et al.* 1961; DE GIER and VAN DEENEN 1961), it was considered of interest to investigate the action of Cl. welchii toxin on the phosphatides of red cells of different animal species.

Methods

About 10 ml of packed red cells, previously washed three times with isotonic saline, were suspended in an equal volume of a solution of 0.9 % sodium chloride and 0.005 M calcium chloride, containing about 300 MLD of a Cl. welchii toxin preparation generously supplied by the Lederle Laboratories, Pearl River N.Y. The mixtures were incubated with gentle shaking at 37°. After the desired incubation periods, red cell ghosts were collected by means of the CO_2 method and the lipids extracted as described before (KÖGL *et al.* 1960). Adequate samples were examined qualitatively by means of "thin-layer chromatography" on silicic acid layers of 0.2 mm (Kieselgel G, Merck). The chromatograms were developed for 2 hrs with

[1]) Contribution No. 18 in the series: Metabolism and functions of phosphatides.

diisobutylketone-acetic acid-water (40:25:5 v/v) (MARINETTI *et al.* 1957). Phosphorus containing components were made visible by spraying with the molybdic acid reagent according to HANES and ISHERWOOD (1949); cephalins and choline containing phosphatides were distinguished by means of the ninhydrine and Dragendorff reagent (KÖGL and VAN DEENEN 1961). It is worth noting that after spraying of the thin-layer chromatograms with the perchloric acid containing molybdic acid reagent the neutal lipids, upon heating, developed brown-black colours.

Quantitative determinations of the lipid and of the phosphatide content, and of the proportions of individual types of phosphatides present in the post-hemolytic residues of red cells, were carried out by procedures described in earlier papers (DE GIER *et al.* 1961; DE GIER and VAN DEENEN 1961).

Results

The effect of Cl. welchii toxin on the phosphatides of red cells appeared to be a very destructive one. Surprisingly, all main types of phosphatides were found to be susceptible to the action of the enzyme, as is shown for the rabbit in fig. 1. The hydrolysis of the phosphatides was accompanied

Fig. 1. Thin-layer chromatogram of lipids from red cell ghosts of rabbit. A. normal. B-D. After incubation with Cl. welchii toxin for 5,15 and 60 min respectively. Sphingomyelins (including lyso-phosphatides), 1; lecithins, 2; cephalins, 3; neutral lipids (cholesterol, glycerides), 4.

by the expected formation of diglycerides, which were recovered in the collected red cell ghosts. Determinations of the phosphatide content of the extractred lipids from ghosts of rabbit red cells revealed that after an incubation of 5 min., which caused a complete hemolysis, already about 42 % of the phosphatides was hydrolysed. Continuation of the incubation during 15 and 60 min. brought about a degradation of the phosphatides of 70 % and 90 % respectively. The proportions of main types of phosphatides we normally found in ghosts of rabbit red cells are: sphingomyelins (including lyso-phosphatides), 26 %; lecithins, 45 %;

cephalins, 29 %. The degrees of hydrolysis of these phosphatides after incubation with Cl. welchii toxin are given below.

sphingomyelins degradation: 5 min., 19 %; 15 min., 40 %; 60 min., 62 %.
lecithins " : " " , 69 %; " " , 90 %; " " , 97 %.
cephalins " : " " , 19 %; " " , 71 %; " " , 92 %.

The different main types of phosphatides present in rabbit red cells and in their post hemolytic residues are apparently all being degradated by the enzyme, although at a rather different rate. Qualitatively the same was found to be true for the action of the bacterial toxin on red cells of ox, pig, man and rat, but in quantitative respect some differences are to be noted. Incubation of e.g. red cells of ox with the toxin for 15 and 60 min. brought about a degradation of 28 % and 66 % of the phosphatides. The phosphatides of ox red cells appeared to consist of: sphingomyelins (including lyso-phosphatides), 59 %; lecithins, 7 %; cephalins, 34 %. As regards the hydrolysis of these phosphatides, the bacterial toxin readily attacked the small amount of lecithins, while the sphingomyelins were degradated faster than the cephalins. Our preliminary results suggest that not only the rates of hydrolysis of various classes of phosphatides are different, but also that the velocity of the enzymatic degradation of a certain type of phosphatide present in red cells or in their post-hemolytic residues may vary between different animal species. Bearing in mind the notable differences in fatty acid patterns of these lipids between red cells of different species (JAMES and LOVELOCK 1957; KOGL *et al.* 1960; HANAHAN *et al.* 1960), it will be of interest to investigate whether these degradation rates are connected to the fatty acid composition of the concerning compounds.

The experiments presented do not allow conclusions about the primary site of the hemolytic action of Cl. welchii toxin, but clearly indicated that the various major types of phosphatides present in red cells are in principle all susceptible to the action of this toxin. The latter fact perhaps explains the findings already reported by OAKLEY and WARRACK (1941), that red cells of various animals are about equally susceptible to this bacterial toxin, although we now know that the phosphatide compositions differ significantly.

Finally, attention has to be paid to the described fact that cephalins in red cell membranes were degradated, whereas isolated, naturally occurring, cephalins (ZAMECKNICK *et al.* 1947; MACFARLANE 1948), as well as synthetic phosphatidyl ethanolamines in an aqueous medium resisted the action of the bacterial toxin (VAN DEENEN *et al.* 1961). This discrepancy was abolished by studying the action of Cl. welchii toxin on synthetic phosphatidyl ethanolamines emulsified together with synthetic lecithins. Although the cephalins alone were not degraded, we observed that in the presence of choline containing phosphatides the phosphatidyl ethanolamines were indeed attacked by Cl. welchii phosphatidase, thereby yielding a component that behaved paper-chromatographically similar to

phosphoryl ethanolamine (phenol-water R_F 0.27). Further experiments on this phenomenon and on the hemolytic action of Cl. welchii toxin are in progress.

Laboratory of Organic Chemistry,
State University Utrecht, The Netherlands

REFERENCES

Deenen, L. L. M. van, G. H. de Haas, C. H. Th. Heemskerk and J. Meduski, Biochem. Biophys. Research Communs., **4**, 183 (1961).
Gier, J. de, I. Mulder and L. L. M. van Deenen, Naturwissenschaften **48**, 54 (1961).
———— and L. L. M, van Deenen, Biochim. Biophys. Acta, **49**, 286 (1961).
Hanahan, D. J., R. M. Watts and D. Pappajohn, J. Lipid Research **1**, 421 (1960),
Hanes, C. S. and F. A. Isherwood, Nature **164**, 1107 (1949).
James, A. T. and J. Lovelock, The blood lipids and the clearing factor. Paleis der Academiën Brussel (1957).
Kögl, F., J. de Gier, I. Mulder and L. L. M. van Deenen, Biochim. Biophys. Acta **43**, 95 1960).
———— and L. L. M. van Deenen, Acta Endocrinologica **36**, 9 (1961).
Macfarlane, M. G., Biochem. J. **42**, 587 (1948).
———— and B. C. J. G. Knight, Biochem. J. **35**, 884 (1941).
Marinetti, G. V., J. Erbland and J. Kochen, Federation Proc. **16**, 837 (1957).
Oakley, C. L. and G. H. Warrack, J. Path. Bact. **53**, 355 (1941).
Zamecknick, P. G., L. E. Brewster and F. Lipman, J. Exptl. Med. **85**, 381 (1947).

DIAGRAM TECHNIQUES IN STATISTICAL MECHANICS. Ia

BY

J. M. J. VAN LEEUWEN

(Communicated by Prof. J. DE BOER at the meeting of April 29, 1961)

§ 1. INTRODUCTION

The recent developments in quantum statistical mechanics are greatly stimulated by the use of diagrams. As the various types of diagrams are almost as numerous as the authors who use them, one might ask whether this great variety is necessary to use diagrams efficiently. The methods used in the theories on the many body problem give the expressions for the properties of the system, as for instance the pressure, as a sum of terms which are weighed integrals over the coordinates of groups of particles. These terms are represented by diagrams for two reasons:

1) to have concise rules for the calculation of the contribution of each term,

2) to have a good administration of the terms occurring.

When a diagram representation is obtained, the main problems are thus: the evaluation of the integral corresponding to each term, and the combinatorial problem of counting the different diagrams.

The main tendency of this article will be to express the properties of the system in terms of "topological" diagrams, i.e. diagrams which are determined by their topological properties only. As will be shown, the use of topological diagrams has many advantages: the corresponding integral expressions and the combinatorial problem are simplified and, moreover, summations over diagrams can be carried out more conveniently with diagrams defined by their topological properties only.

In classical statistical mechanics the topological properties of the classical (Mayer) diagrams are already the only essential ones, because they determine the value of the corresponding integral as well as the combinatorial problem.

In quantum statistical mechanics the situation is more complicated. In most of the diagram representations many terms are represented by different diagrams, distinguished from each other by non-topological features. It is the aim of this work to show that these non-topological features can be removed in a convenient way for some types of diagrams frequently used [1, 2]. Starting from the diagrams introduced by LEE and YANG [1] for the statistical operator $\exp -\beta\mathcal{H}$ we arrive, via a stage

which resembles the diagrams used by Bloch and de Dominicis [2], at a diagram representation of the grand canonical partition function and the reduced density matrices in which only the topological properties of the diagrams are relevant. As far as the grand canonical partition function Z_{gr} is concerned the obtained diagrams are identical to those introduced by Montroll and Ward [3].

In subsequent papers it will be shown that this diagram expansion is very well suited to the application to hard core potentials and to the degenerate Bose gas.

Preceding the proper derivation of the "topological" diagrams in § 4 we give in § 1 some definitions in order to define the notation and to make it possible to treat the three types of statistics simultaneously: Bose–Einstein statistics, Boltzmann statistics and Fermi–Dirac statistics.

In § 2 we give the formulae for the diagram representation for the partition function and the reduced density matrices in terms of diagrams as defined by Lee and Yang [1] for the statistical operator $\exp -\beta\mathscr{H}$.

§ 3 contains the usual reduction to the connected diagrams both for the reduced density matrices as well as for the partition function. It is not necessary to do this before the transition to "topological" diagrams, but it makes the discussion of the latter easier. It turns out that the expansion in terms of connected diagrams is more readily obtained than the corresponding one for the partition function. This is a general feature of manipulations with diagrams, which finds its origin in the fact that the presence of reference particles supresses certain symmetries which complicate the diagram representation for the partition function.

Therefore, first the "topological" diagrams are derived for the reduced density matrices (§ 4), by successively eleminating the non-topological features in the diagrams.

The treatment finishes with the discussion of the "topological" diagrams for the partition function (§ 6).

We shall be concerned in this paper with the equilibrium properties of systems consisting of a large number N of identical particles in a large volume V. Each particle i has the coordinates q_i being the position coordinate r_i and the internal variables, which we restrict to a spin coordinate. The complete configuration space of the system will be denoted by $q^N=(q_1 \ldots q_N)$. The system is characterised by its Hamilton operator $\mathscr{H}_N$. This Hamiltonian determines together with the boundary condition the states inside the volume V. For convenience the boundary condition is taken to be periodic. $\mathscr{H}_N$ is assumed to be of the form:

$$\mathscr{H}_N = \mathscr{H}_N{}^0 + \Phi_N \tag{1}$$

in which $\mathscr{H}_N{}^0$ represents the Hamiltonian of the ideal gas of N particles and Φ_N accounts for their interaction. $\mathscr{H}_N{}^0$ is a sum of N 1-particle operators and Φ_N is the sum of $N(N-1)/2$ 2-particle operators being the interaction potentials between the $N(N-1)/2$ pairs of particles. In all

applications the potentials are assumed to depend only on the relative distances r_{ij} of the particles. So:

$$\Phi_N = \sum_{i<j} \phi(r_{ij}). \tag{2}$$

The N-particle states may be expanded in terms of the N-particle eigenstates $|k^N\rangle$ of the unperturbed Hamiltonian $\mathscr{H}_N{}^0$. These eigenstates $|k^N\rangle$ are products of 1-particle eigenstates $|k_i\rangle$. The k_i are the quantum numbers that specify the 1-particle states in the volume V. k_i contains the three components of the momentum vector and an internal quantum number for the spin states.

We must select properly symmetrised states according to the type of statistics. Throughout these papers we distinguish the three types of statistics by means of a label ε namely as follows:

$$\begin{cases} \varepsilon = \ \ 0 \text{ Boltzmann statistics} \\ \varepsilon = \ \ 1 \text{ Bose–Einstein statistics} \\ \varepsilon = -1 \text{ Fermi–Dirac statistics.} \end{cases} \tag{3}$$

In forming symmetrised states lineair combinations are taken of the permuted states: $|\mathrm{P}k^N\rangle$

$$|\mathrm{P}k^N\rangle = |\mathrm{P}(k_1 \ldots k_N)\rangle = |k_{\mathrm{P}(1)} \ldots k_{\mathrm{P}(N)}\rangle$$

where $\mathrm{P}(1), \mathrm{P}(2) \ldots \mathrm{P}(N)$ is a permutation of the numbers $1, 2 \ldots N$. The selection of the properly symmetrised states is made, by use of the symmetrisation operator:

$$\mathscr{S}_\varepsilon = \frac{1}{N!} \sum_{k^N} \sum_{\mathrm{P}} \varepsilon^{N-c_\mathrm{P}} |k^N\rangle\langle \mathrm{P}k^N| \tag{4}$$

c_P is the number of cycles in the permutation P. For Boltzmann statistics ($\varepsilon = 0$) only the identical permutation contributes ($c_\mathrm{P} = N$). So $\mathscr{S}_0$ is $1/N!$ times the unit operator. We point out that for $\varepsilon = \pm 1$, $\mathscr{S}_\varepsilon$ is a projection operator i.e.:

$$\mathscr{S}^2_{\pm 1} = \mathscr{S}_{\pm 1}.$$

Also $\mathscr{S}_\varepsilon$ commutes with all operators which are symmetric in the N particle; for instance:

$$\mathscr{S}_\varepsilon \mathscr{H}_N = \mathscr{H}_N \mathscr{S}_\varepsilon.$$

We will use both the *canonical ensemble* and the *grand canonical ensemble*. For the canonical ensemble the statistical operator reads:

$$\mathscr{W}_N(\beta) = \exp - \beta \mathscr{H}_N \qquad \beta = 1/kT \tag{5}$$

which is defined by the power series expansion. In terms of the statistical operator $\mathscr{W}_N(\beta)$ are defined:

a) The partition function Z_N which is the trace of the (symmetrised) statistical operator:

$$Z_N = \mathrm{tr}\ \mathscr{S}_\varepsilon \mathscr{W}_N(\beta). \tag{6}$$

b) The reduced density matrices $n_h^{(N)}(q'^h, q^h)$ $(h>1)$:

$$n_h^{(N)}(q'^h, q^h) = \frac{N!}{(N-h)!}\frac{1}{Z_N}\int dq_{h+1}\ldots dq_N \langle q'^h, q_{h+1}\ldots q_N | S_\varepsilon \mathscr{W}_N(\beta)| q^N\rangle. \tag{7}$$

The diagonal elements of these reduced density matrices have a direct physical meaning namely:

$$\begin{cases} h=1 & n_h^{(N)}(q_1, q_1) = \text{particle density} \\ h=2 & n_h^{(N)}(q_1 q_2, q_1 q_2) = \text{pair distribution function, etc.} \end{cases} \tag{8}$$

The factor $1/N!$ in the definition of $\mathscr{S}_0$ for the Boltzmann statistics provides in (6) and (7) that the results for the Boltzmann statistics are the high temperature limits of the two other types of statistics. We will set up our treatment in such a way that it applies simultaneously to the three types of statistics. Therefore we use the ε as a parameter (as in (4)) and restrict afterwards ε to the values ± 1 and 0. Of course this is rather weighty for the Boltzmann statistics since the simplifications come in from the beginning when $\varepsilon=0$. We give up these simplifications deliberately as our main interest goes out to the other types of statistics and as all Boltzmann results can be obtained without too much difficulty by taking afterwards $\varepsilon=0$.

The grand canonical partition function Z_{gr} is related to Z_N by:

$$Z_{\mathrm{gr}} = \sum_N \bar{z}^N Z_N = \sum_N \exp(\beta\mu N) Z_N \tag{9}$$

where the constant μ is the thermodynamical potential per molecule which can be fixed by the equation:

$$\bar{z} = \exp\beta\mu = \lim_{N\to\infty}\frac{Z_{N-1}}{Z_N}. \tag{10}$$

§ 2. Diagram representations for the partition function and the reduced density matrices

In this section the operator $\mathscr{W}_N(\beta)$ will be decomposed into ascending powers of Φ_N. The terms occurring can be characterised by diagrams. The same type of diagrams are used in the study of the contributions to the partition function Z_N and the reduced density matrices $n_h^{(N)}(q'^h, q^h)$.

The expansion of the statistical operator $\mathscr{W}_N(\beta)$ into powers of Φ_N results by iteration of the operator equation for $\mathscr{W}_N(\beta)$:

$$\mathscr{W}_N(\beta) = \mathscr{W}_N{}^0(\beta) - \int_0^\beta \mathscr{W}_N{}^0(\beta-\beta_1)\,\Phi_N\,\mathscr{W}_N(\beta_1)\,d\beta_1 \tag{11}$$

in which $\mathscr{W}_N^0(\beta)$ is the unperturbed statistical operator $\exp -\beta \mathscr{H}_N^0$. The iteration gives:

$$(12)\quad \begin{cases} \mathscr{W}_N(\beta) = \mathscr{W}_N^0(\beta) + \int_0^\beta \mathscr{W}_N^0(\beta-\beta_1)\,(-\Phi_N)\,\mathscr{W}^0(\beta_1)\,d\beta_1 + \\ + \int\limits_{\beta>\beta_1>\beta_2>0} d\beta_1\,d\beta_2\,\mathscr{W}_N^0(\beta-\beta_1)\,(-\Phi_N)\,\mathscr{W}_N^0(\beta_1-\beta_2)\,(-\Phi_N)\,\mathscr{W}_N^0(\beta_2)\ldots \end{cases}$$

The term of order p in Φ_N consists itself out of $[N(N-1)/2]^p$ terms arising from the $N(N-1)/2$ terms in Φ_N. Following LEE and YANG [1] the various terms can be represented by means of (ordered) diagrams. To illustrate this we pick out some typical term out of the third order ($p=3$) in Φ_N (with $N=5$)

$$(13)\quad \int\limits_{\beta>\beta_1>\beta_2>\beta_3>0} d\beta_1\,d\beta_2\,d\beta_3\;\mathscr{W}_5^0(\beta-\beta_1)\,(-\phi_{13})\,\mathscr{W}_5^0(\beta_1-\beta_2)\,(-\phi_{24})\cdot \\ \cdot\,\mathscr{W}_5^0(\beta_2-\beta_3)\,(-\phi_{34})\,\mathscr{W}_5^0(\beta_3).$$

The diagram corresponding to this term is drawn in fig. 1. It consists of N ($=5$) ordered vertical (particle) lines with the numbers $1, \ldots N$ and p ($=3$) horizontal ordered (interaction) lines with the numbers $1, \ldots p$. The interaction lines connect the particle lines as given by of the interactions of the corresponding term (13). The sequence of the interactions from the lower side to the upper side in the diagram is the same as the order from right to left in the represented term (13). Each interaction is associated with an inverse temperature β_s. This gives a 1:1 correspondence of these *ordered diagrams* and terms in the expansion (12).

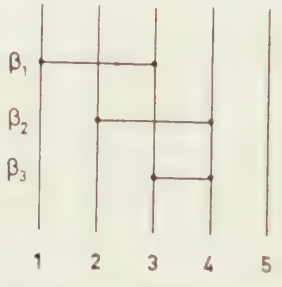

Fig. 1. Diagram for the statistical operator $\mathscr{W}_N(\beta)$ (c.f. (13)).

The partition function Z_N and the reduced density matrices $n_h^{(N)}(q'^h, q^h)$ can be expressed in terms of the matrix elements $\langle k'^N|\mathscr{W}_N(\beta)\,k^N\rangle$. We discuss therefore first the contributions of the diagrams to $\langle k'^N|\mathscr{W}_N(\beta)|k^N\rangle$. It is preferable to remove all operators out of the diagrams for

$$\langle k'^N|\mathscr{W}_N(\beta)|k^N\rangle.$$

This can be done by writing for the intermediate operators:

$$(14)\quad \begin{cases} \mathscr{W}_N^0(\beta_s-\beta_{s+1}) = \\ \quad = \sum_{p_1^s \ldots p_N^s} |p_1^s \ldots p_N^s\rangle \exp\left[-(\beta_s-\beta_{s+1})\,E(p_1^s \ldots p_N^s)\right] \langle p_1^s \ldots p_N^s| \end{cases}$$

where $E(p_1^s \ldots p_N^s)$ is the unperturbed energy of the intermediate state $|p_1^s \ldots p_N^s\rangle$ between the interactions $s+1$ and s. It is the sum of 1-particle energies $E(p_j^s)$. So:

$$(15)\quad \begin{cases} \exp\left[-(\beta_s-\beta_{s+1})\, E(p_1^s \ldots p_N^s\right] = \\ \qquad = \exp\left[-(\beta_s-\beta_{s+1}) \sum_{j=1}^{N} E(p_j^s)\right] = \prod_{j=1}^{N} \langle p_j^s | \mathscr{W}_1^0(\beta_s-\beta_{s+1}) | p_j^s \rangle . \end{cases}$$

Working out all bra-vectors $\langle p_1^s \ldots p_N^s|$ and ket-vectors $|p_1^s \ldots p_N^s\rangle$ the final expression $\langle k'^N | \mathscr{W}_N(\beta) | k^N \rangle$ contains only the matrix elements

$$\langle p_i^s p_j^s | -\phi | p_i^{s+1} p_j^{s+1} \rangle$$

for the interaction lines and the particle propagators

$$\langle p_j^s | \mathscr{W}_1^0(\beta_s-\beta_t) | p_j^s \rangle = \exp -(\beta_s-\beta_t)\, E(p_j^s)$$

for the particles lines (β_s and β_t are successive interactions on the j-th particle line).

As diagram for $\langle k'^N | \mathscr{W}_N(\beta) | k^N \rangle$ we can take the same figure as the diagram of e.g. fig. 1 for the operator $\mathscr{W}_N(\beta)$, by adding to all particle line segments the corresponding quantum numbers:

for the internal segments: p_j^s
for the upper loose ends : k_j'
for the lower loose ends : k_j

(j is the number of the particle line and s the number of the interaction at the upper end of the segment as shown in fig. 2).

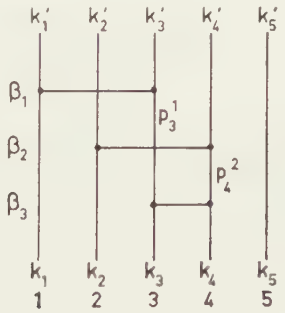

Fig. 2. Diagram for $\langle k'^5 | \mathscr{W}_5(\beta) | k^5 \rangle$ (c.f. (18)).

To each line of such a diagram a factor is associated in the following way:
Firstly:

$$f'_{\text{segm}} = \langle p_j^s | \mathscr{W}_1^0(\beta_s-\beta_t) | p_j^s \rangle \qquad \text{(i')}$$

to the *segment* on the j-th particle line between the interaction t and s. For an upper loose end we must set $\beta_s = \beta$ and $p_j^s = k_j'$; for a lower loose

end $\beta_t = 0$ and $p_j^s = k_j$; for a particle line without interaction (i') becomes:

$$\delta(k_j', k_j)\, \langle k_j | \mathscr{W}^0(\beta) | k_j \rangle .$$

Secondly:

$$f'_{\text{int}} = \langle p_j^s\, p_{j'}^{s} | -\phi | p_j^r\, p_{j'}^{r'} \rangle \tag{ii'}$$

to the s-th *interactionline* with the appropriate changes of the p_j^s into k_j' (k_j) for the first (last) interaction on the j-th particle line. Then we have:

$$\langle k'^N | \mathscr{W}_N(\beta) | k^N \rangle = \sum_{\Gamma_N} C'(\Gamma_N) \tag{16}$$

where the summation runs over all ordered diagrams Γ_N and the contribution $C'(\Gamma_N)$ of each of the ordered diagrams Γ_N is calculated as:

$$C'(\Gamma_N) = \int\limits_{\beta > \beta_1 \ldots > \beta_p > 0} d\beta_1 \ldots d\beta_p \sum_{\{p_j^s\}} \left(\prod f'_{\text{segm}}\, f'_{\text{int}} \right)_{\Gamma_N} . \tag{17}$$

The product is extended over all particle line segments and interaction lines of the diagram Γ_N. As an example we give the contribution of the diagram drawn in fig. 2.

$$\begin{cases} \displaystyle\int\limits_{\beta>\beta_1>\beta_2>\beta_3>0} d\beta_1 d\beta_2 d\beta_3 \sum_{p_3', p_4^2} \langle k_1' | \mathscr{W}^0(\beta-\beta_1) | k_1' \rangle \langle k_2' | \mathscr{W}^0(\beta-\beta_2) | k_2' \rangle \\ \times \langle k_3' | \mathscr{W}^0(\beta-\beta_1) | k_3' \rangle \langle k_4' | \mathscr{W}^0(\beta-\beta_2) | k'_4 \rangle \, \delta(k_5', k_5) \, \langle k_5 | \mathscr{W}^0(\beta) | k_5 \rangle \\ \times \langle k_1' k_3' | -\phi | k_1 p_3^1 \rangle \langle k_1 | \mathscr{W}^0(\beta_1) k_1 \rangle \langle p_3^1 | \mathscr{W}^0(\beta_1 - \beta_3) | p_3^1 \rangle \\ \times \langle k_2' k_4' | -\phi | k_2 p_4^2 \rangle \langle k_2 | \mathscr{W}^0(\beta_2) | k_2 \rangle \langle p_4^2 | \mathscr{W}^0(\beta_2 - \beta_3) | p_4^2 \rangle \\ \times \langle p_3^1 p_4^2 | -\phi | k_3 k_4 \rangle \langle k_3 | \mathscr{W}^0(\beta_3) | k_3 \rangle \langle k_4 | \mathscr{W}^0(\beta_4) | k_4 \rangle . \end{cases} \tag{18}$$

Although expressions like (18) are lenghty and not transparant, they show already the advantage of the diagram representation, which lies in the fact that the contribution of each diagram is completely expressible in the 1-particle functions $\langle p_j^s | \mathscr{W}^0(\beta_s - \beta_t) | p_j^s \rangle$ and the 2-particle functions $\langle p_j^s p_{j'}^{s} | -\phi | p_j^r p_{j'}^{r} \rangle$ and is, in principle, computable by integrations and summations.

Once having a diagram representation for the matrix elements $\langle k'^N | \mathscr{W}^N(\beta) | k^N \rangle$, a similar representation in diagrams can be obtained for the partition function and for the reduced density matrices by making a few extensions. From (6) we have:

$$\begin{cases} Z_N = \dfrac{1}{N!} \displaystyle\sum_{k'^N} \sum_{k^N} \sum_{\mathrm{P}} \varepsilon^{N-C_\mathrm{P}} \langle k^N | k'^N \rangle \langle \mathrm{P} k'^N | \mathscr{W}^N(\beta) | k^N \rangle = \\ \qquad = \dfrac{1}{N!} \displaystyle\sum_{k^N} \sum_{\mathrm{P}} \varepsilon^{N-C_\mathrm{P}} \langle \mathrm{P} k^N | \mathscr{W}^N(\beta) | k^N \rangle . \end{cases} \tag{19}$$

Starting from the previous diagrams, the row $k_1' \ldots k_N'$ at the top of the diagram is replaced by the row $k'_{\mathrm{P}(1)} \ldots k'_{\mathrm{P}(N)}$. For the *partition function* Z_N it is necessary to take

$$k'_{\mathrm{P}(j)} = k_{\mathrm{P}(j)} \qquad \text{for all } j.$$

These equalities are expressed in the diagram [2] by connecting the top of the j-th particle line with the lowerside of the particle line $\mathrm{P}(j)$ by a dotted line. This gives also a representation of the permutation P in the diagram. In the diagrams extended in this way the row at the top $k'_{\mathrm{P}(1)} \ldots k'_{\mathrm{P}(N)}$ may then be dropped. The factor $\varepsilon^{N-c_{\mathrm{P}}}$ can be read directly from the extended diagrams since the number of cycles c_{P} in the permutation P equals the number of closed loops in the particle and dotted lines. We will call these extended diagrams the $(0, N)$ *diagrams* and we have the following result:

$$Z_N = \frac{1}{N!} \sum_{\Gamma_{0N}} C'(\Gamma_{0N}) \tag{19a}$$

where the contribution $C'(\Gamma_{0N})$ of each of the $(0, N)$ diagrams Γ_{0N} is calculated as

$$C'(\Gamma_{0N}) = \int_{\beta > \beta_1 > \ldots > \beta_p > 0} d\beta_1 \ldots d\beta_p \sum_{k^N} \sum_{\{p_j{}^s\}} (\varepsilon^{N-c_{\mathrm{P}}} \prod f'_{\mathrm{segm}} f'_{\mathrm{int}})_{\Gamma_{0N}}. \tag{20}$$

In order to obtain a diagram representation for the *reduced density matrices* one has from (7):

$$\left.\begin{aligned} n_h{}^{(N)}(q'^h, q^h) = \frac{1}{(N-h)!} \int dq_{h+1} \ldots dq_N \sum_{k'^N} \sum_{k^N} \varepsilon^{N-c_{\mathrm{P}}} \\ \cdot \langle q'^h, q_{h+1} \ldots q_N | k'^N \rangle \langle \mathrm{P}k'^N | \mathscr{W}_N(\beta) | k^N \rangle \langle k^N | q^N \rangle. \end{aligned}\right\} \tag{21}$$

Taking the diagrams for $\langle k'^N | \mathscr{W}_N(\beta) | k^N \rangle$ again as starting point we have to change the top row $k_1' \ldots k_N'$ into $k'_{\mathrm{P}(1)} \ldots k'_{\mathrm{P}(N)}$. Integrating over the coordinates of the particles $h+1, \ldots N$ leads only to non-zero contributions when:

$$k'_{\mathrm{P}(j)} = k_{\mathrm{P}(j)} \qquad \text{for } j \text{ with } h < \mathrm{P}(j) \leqslant N.$$

These conditions on the final states can be visualized by again connecting the top of the j-th particle line to the lower end of the particle line $\mathrm{P}(j)$ for all $\mathrm{P}(j) > h$. In this way the permutation P is only partially indicated in the diagram. It can be completed by writing the number $\mathrm{P}(j)$ at the open ends at the top of the j-th particle line. The quantum numbers k_j' for the closed particle lines at the top can be omitted. The numbers $k_1' \ldots k_h'$ are left. We call these type of extended diagrams the (h, N) *diagrams*. Then we have:

$$n_h{}^{(N)}(q'^h, q^h) = \frac{1}{(N-h)!\, Z_N} \sum_{k'^h} \sum_{k^h} \langle q'^h | k'^h \rangle \Big[\sum_{\Gamma_{hN}} C'(\Gamma_{hN}) \Big] \langle k^h | q^h \rangle \tag{22}$$

where the contribution $C'(\Gamma_{hN})$ of each of the (h, N) diagrams Γ_{hN} is calculated as:

$$C'(\Gamma_{hN}) = \int_{\beta > \beta_1 > \ldots \beta_p > 0} d\beta_1 \ldots d\beta_p \sum_{k_{h+1} \ldots k_N} \sum_{\{p_j{}^s\}} (\varepsilon^{N-c_{\mathrm{P}}} \prod f'_{\mathrm{segm}} f'_{\mathrm{int}})_{\Gamma_{hN}}. \tag{23}$$

The number of closed loops c_P must be calculated as if *all* upper ends were connected to the lower ends according to the permutation P. In order to illustrate the procedure in fig. 3 we have drawn extensions of the diagram of fig. 1 according to the permutation P(12345)=(24351).

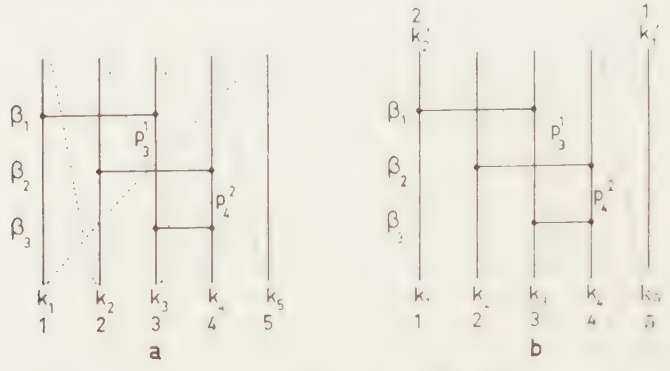

Fig. 3. Diagrams for the partition function Z_5 (3a) and for the reduced density matrix $n_2^{(5)}(q'^2, q^2)$ (3b c.f. (24)).

In fig. 3a the extended version as (0,5) diagram contributing to Z_N is given; in fig. 3b the version which contributes as (2, 5) diagram to $n_2^{(5)}(q_1'q_2', q_1q_2)$. The contribution of the diagram of fig. 3b reads:

$$
(24)\quad \begin{cases} \int\limits_{\beta>\beta_1>\beta_2>\beta_3>0} d\beta_1\, d\beta_2\, d\beta_3 \sum_{k_3 k_4 k_5} \sum_{p_3{}^1 p_4{}^2} \langle k_2'|\mathscr{W}^1{}^0(\beta-\beta_1)|k_2' \\ \langle k_4|\mathscr{W}^1{}^0(\beta-\beta_2+\beta_3)|k_4\rangle \langle k_3|\mathscr{W}^1{}^0(\beta-\beta_1+\beta_3)|k_3\rangle \langle k_5|\mathscr{W}^1{}^0(2\beta-\beta_2)|k_5\rangle \\ \delta(k_1', k_5)\, \langle k_2' k_3|-\phi|k_1 p_3{}^1\rangle \langle k_1|\mathscr{W}^1{}^0(\beta_1)|k_1\rangle \langle p_3{}^1|\mathscr{W}^1{}^0(\beta_1-\beta_3)|p_3{}^1\rangle \\ \langle k_4 k_5|-\phi|k_2 p_4{}^2\rangle \langle k_2|\mathscr{W}^1{}^0(\beta_2)|k_2\rangle \langle p_4{}^2|\mathscr{W}^1{}^0(\beta_2-\beta_3)|p_4{}^2\rangle \langle p_3{}^1 p_4{}^2|-\phi|k_3 k_4\rangle. \end{cases}
$$

Although the form of the (h, N) diagrams is provisional (in § 4 they get their final form), it is not merely a matter of taste. In the following section it turns out that the dotted lines play an essential role in the connectiveness of the diagram. In § 4 it becomes clear that the pairs of loose ends connected by a dotted line are indeed equivalent in all respects with the internal lines.

§ 3. Reduction to connected diagrams

Until now the diagrams all involved N particle lines. For convenience only we have chosen $N = 5$ in our examples (figs. 1, 2 and 3), as it would be difficult to visualize the diagrams with N particle lines. However, it is possible to factorize the contributions of the (most important) N particle diagrams, in contributions of diagrams with a smaller number of particle lines, in complete analogy with the classical diagrams [4], the factorisation of which leads to the Ursell expansion [5]. This reduction is based on the concept of *connected* diagrams.

A $(0, N)$ diagram is connected if every particle line is connected to

the others directly or indirectly by interaction lines or dotted lines. In the (h, N) diagrams $(h>0)$ the first h particle lines play a special role, and such a (h, N) diagram is said to be "connected" if the other particle lines are connected to at least one of the first h particle lines.

In order to write the contributions of the diagrams as products of the contributions of the constituting connected parts, we have to carry out a summation of all those diagrams, which differ only in the order of the interaction lines of one disconnected part with respect to those of another disconnected part (figs. 4a, b). Due to the fact that the integrand of (20)

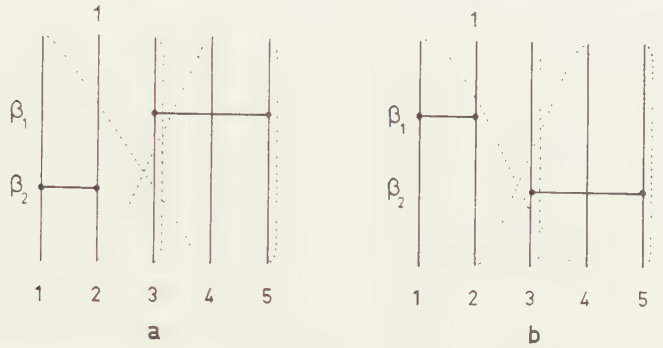

Fig. 4. Non-connected diagrams. The subdiagram consisting of the particle lines 3 and 5 is not connected to the part consisting of the particle line 1.

is the product of the integrands corresponding to the disconnected parts (see e.g. figs. 4a and b), this summation can be performed by integrating (20) over the β_{s_i} and the β_{t_i} of the two disconnected parts respectively, i.e. under the conditions:

$$\beta > \beta_{s_1} > \ldots > 0$$
$$\beta > \beta_{t_1} > \ldots > 0.$$

Then the whole expression (20) can be divided into two parts, each involving only factors, summations and integrations coming from one of the disconnected parts. Therefore, taking together the diagrams which differ only in the relative order of the interaction lines out of the disconnected parts, the contribution of such a set is the product of the contributions of the disconnected parts.

By using this factorisation property, the disconnected parts can be separated. In doing this the arguments for the $(0, N)$ diagrams contributing to the partition function and for the (h, N) diagrams contributing to the reduced density matrices are somewhat different. Since the latter case is easier we shall treat this first.

A) *The reduced density matrices* $(h>0)$

Consider a (h, N) diagram with a connected part consisting of the first h particle lines and l additional particle lines connected to these

first h particle lines, and a remaining part which is not connected to the first h particle lines. The contribution corresponding to all possible connected diagrams consisting of $h+l$ particle lines will be represented by:

$$l!\, a_l^{(h)}(q'^h, q^h) = \sum_{k'^h} \sum_{k^h} \langle q'^h k^h \rangle \Big[\sum_{\Gamma^c_{h,h+l}} C'(\Gamma^c_{h,h+l}) \Big] \langle k^h | q^h \tag{25}$$

where the contribution $C'(\Gamma^c_{h,h+l})$ of each connected (h, N) diagram $\Gamma^c_{h,h+l}$ is defined by (23). The contribution of all possible remaining parts consisting of $N-h-l$ particle lines is represented by:

$$(N-h-l)!\, Z_{N-h-l}.$$

As there are $\binom{N-h}{N-h-l}$ possible ways of taking out of the $N-h$ molecules $h+1, \ldots N$ a group of $N-h-l$ molecules, the total contribution resulting from all (h, N) diagrams with a connected part of $h+l$ particle lines is given by:

$$\binom{N-h}{N-h-l} [l!\, a_l^{(h)}(q'^h, q^h)]\, [N-h-l)!\, Z_{N-h-l}].$$

Carrying out the summation over all values of l gives:

$$n_h^{(N)}(q'^h, q^h) = \frac{1}{(N-h)!\, Z_N} \sum_{l=0}^{N-h} \binom{N-h}{N-h-l} [l!\, a_l^{(h)}(q'^h, q^h)]\, [(N-h-l)!\, Z_{N-h-l}]$$

or:

$$n_h^{(N)}(q'^h, q^h) = \sum_{l=0}^{N-h} a_l^{(h)}(q'^h, q^h)\, Z_{N-h-l}/Z_N. \tag{26}$$

Using (10) $h+l$ times we have

$$\lim_{N\to\infty} Z_{N-h-l}/Z_N = \bar{z}^l$$

so that by an inter-change of sum and limit (26) is transformed into:

$$n_h(q'^h, q^h) = \lim_{N\to\infty} n_h^{(N)}(q'^h, q^h) = \sum_{l=0}^{\infty} a_l^{(h)}(q'^h, q^h)\, \bar{z}^{h+l}. \tag{27}$$

The limit function $n_h(q'^h, q^h)$ is usually taken as *the* reduced density matrix of h particles. It is thus by (27) expressed in the connected $(h, h+l)$ diagrams only.

B) *The partition function* $(h=0)$

The $(0, N)$ diagrams cannot be divided unambiguously like the (h, N) diagrams into a connected part and a remaining disconnected part. In general a $(0, N)$ diagram breaks down into connected parts with $l_1, l_2 \ldots$ particle lines. Again we define:

$$l!\, a_l^{(0)} = \sum_{\Gamma^c_{0l}} C'(\Gamma^c_{0l}) \tag{28}$$

where $C'(\Gamma_{0l}^{c})$ is the contribution of the connected $(0, N)$ diagram Γ_{0l}^{c} given by (23). If m_l is the number of connected components with l particle lines, then all $(0, N)$ diagrams with a partition $\{m_l\}$ contribute:

$$\prod_l \frac{N!}{[l!]^{m_l}\, m_l!}\, [l!\, a_l^{(0)}]^{m_l} \qquad \sum l m_l = N.$$

The factor $m_l!$ in the denominator of the combinatorial factor corrects for distribution of the connected parts which do not lead to new diagrams [2]. For the contribution of all $(0, N)$ diagrams we have to total over all partitions $\{m_l\}$ under the restriction $\sum l m_l = N$:

$$Z_N = \sum_{\{m_l\}}{}' \prod_l \frac{[a_l^{(0)}]^{m_l}}{m_l!} \qquad \sum_l l m_l = N.$$

This formula is the same as in the classical theory [5] because $a_l^{(0)} = V b_l/\lambda^{3l}$ (b_l is the l-th cluster integral and λ is the thermal wavelength). We can get rid of the restriction (indicated by the prime) on the summation over the partitions $\{m_l\}$ by going over to the grand canonical partition function Z_{gr}:

$$Z_{\mathrm{gr}} = \sum_N \bar{z}^N Z_N = \prod_l \sum_{\{m_l\}} \frac{[\bar{z}^l\, a_l^{(0)}]^{m_l}}{m_l!} = \exp \sum_{l=1}^{\infty} a_l^{(0)}\, \bar{z}^l$$

or:

$$\beta p V = \ln Z_{\mathrm{gr}} = \sum_{l=1}^{\infty} a_l^{(0)}\, \bar{z}^l. \tag{29}$$

As it is easy to see that the contribution of a connected diagram is proportional to the volume V, the reduction to the connected diagrams is also a selection of the contributions according to their dependence on the volume V. We thus see that the reduction to the connected diagrams runs parallel with the (generalized) Ursell expansion. In fact the functions $a_l^{(h)}(q'^h, q^h)$ are related to the cluster functions $b_l^{(h)}(q'^h, q^h)$ introduced by DE BOER [5] according to:

$$(l+1)\, b_{l+1}^{(h)}(q'^h, q^h) = \lambda^{3(h+l)}\, a_l^{(h)}(q'^h, q^h). \tag{30}$$

The transition from the canonical ensemble to the grand canonical ensemble introduces the new variable $\bar{z}$ (or the thermodynamical potential μ) instead of the number of particles N. The elimination of $\bar{z}$ goes with the aid of the relation:

$$N = \int n_1(q_1, q_1)\, dq_1 = \sum_{l=0}^{\infty} \bar{z}^{l+1} \int a_l^{(1)}(q_1, q_1)\, dq_1. \tag{31}$$

For the righthand side we can use the relation:

$$\int l!\, a_l^{(1)}(q_1, q_1)\, dq_1 = (l+1)!\, a_{l+1}^{(0)} \tag{32}$$

which follows by integration of the $(1, l)$ diagrams (contributing to the

left-hand side) over the coordinate $q_1' = q_1$ giving the $(0, l+1)$ diagrams (contributing to the right-hand side).

This shows that (29) and (31) are in agreement with the thermodynamical relation:

$$N = \sum_{l=1} l a_l^{(0)} \bar{z}^l = \left(\frac{\partial pV}{\partial \mu}\right)_{T,V}. \tag{33}$$

In the fluid state the functions $a_l^{(1)}(q_1, q_1)$ are independent of the coordinate q_1. So with (30) and (32) we have:

$$a_l^{(0)} = V a_{l-1}^{(1)}/l = V\lambda^{-3l} b_l^{(1)} \stackrel{\text{def}}{=} V\lambda^{-3l} b_l$$

Then (33) and (29) can be written in the more usual form:

$$n = N/V = \sum_{l=1}^{\infty} l b_l z^l \qquad z = \bar{z}/\lambda^3 = e^{\beta\mu}/\lambda^3. \tag{34}$$

$$\beta p = \sum_{l=1}^{\infty} b_l z^l \tag{35}$$

(To be continued)

DIAGRAM TECHNIQUES IN STATISTICAL MECHANICS. Ib

BY

J. M. J. VAN LEEUWEN

(Communicated by Prof. J. DE BOER at the meeting of April 29, 1961)

§ 4. INTRODUCTION OF "TOPOLOGICAL" DIAGRAMS FOR THE REDUCED DENSITY MATRICES

The aim of this section is to derive rules for the calculation of the contribution of all diagrams of the same topological type. The topological structure of a diagram is characterised by the type of connection of the *toron lines* by means of the interaction lines. A *toron line* [3] is a chain of particle lines connected to each other by dotted lines. There are in a $(h, h+l)$ diagram h open toron lines, starting with the first h particle lines. These open toron lines may be called reference lines in analogy with the classical case [4]. The other toron lines are closed.

As the case of the $(0, l)$ diagrams has again some complications we deal in this section only with the $(h, h+l)$ diagrams and postpone the treatment of the $(0, l)$ diagrams to the next section.

We start with a discussion of the possible non-topological features.

A) *Non-topological features*

(1) *The order in which the last l particle lines occur* in the diagram is, as such, a non-topological feature. (The first h particle lines have as starting-points of the h reference lines a special position in the diagram). Diagrams differing only by a permutation of these particle lines (together with the end-points of the interaction lines ending on these particle lines) can be obtained by a permutation of the molecules only and thus have the same contribution. This class of $l!$ diagrams (all $l!$ are different !) may be represented by one of them, the contribution of which then should be multiplied by $l!$. This factor $l!$ cancels out against the $l!$ occurring in the definition (25) of $a_l^{(h)}(q'^h, q^h)$. So the elimination of this non-topological feature leads to the cancellation of the factor $l!$ in (25).

(2) A second non-topological feature lies in *the order of some interaction lines.* To illustrate this, it is convenient to draw the diagrams on a cylinder [2] with the circumference β; the dotted lines are then running at the rear of the cylinder. The direction of increasing β is indicated by an arrow in the particle lines. In figs. 5a, b, c and d examples are drawn of (1, 5) diagrams of the same topological structure. We have used the interchangeability of the particle lines to draw the successive windings

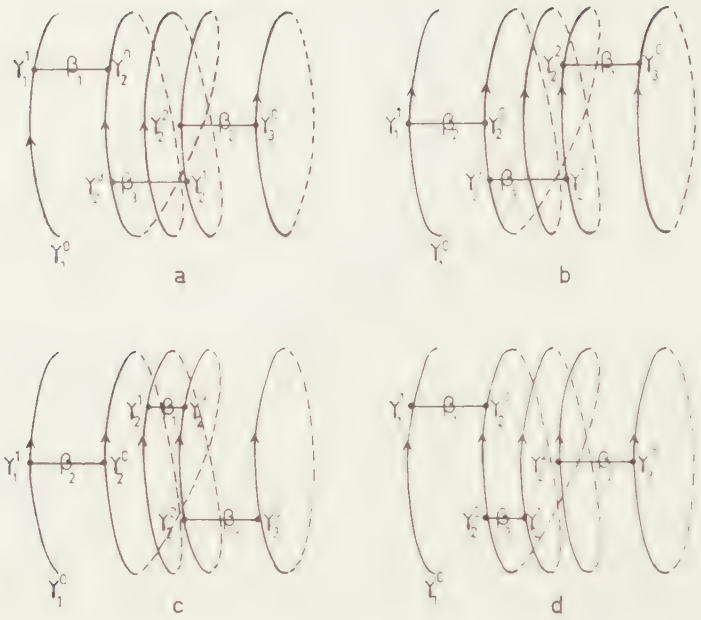

Fig. 5. Topologically equivalent diagrams as discussed under (2).

of a toron line next to each other. In fig. 5, diagram b) differs from diagram a) in the relative position of the interaction lines β_1 and β_2 (by convention this numbering is always in the direction of decreasing β). Diagram c) results from diagram a) by moving interaction β_3 of a) parallel to itself at the rear of the cylinder in the direction of decreasing β till it reaches the position of interaction β_1 in diagram c). Also d) follows from a) by moving the righthand side vertex of interaction β_3 of a) a complete winding in the direction of decreasing β.

(3) Thirdly, two diagrams are topologically equivalent when they differ only in *the number of windings per toron line*. As an example figs. 6a and 6c show (1, 4) diagrams which are topologically equivalent to the diagrams 5a and 5c by this argument. They can be obtained from the diagrams 5a and 5c by omitting a winding in the second toron line.

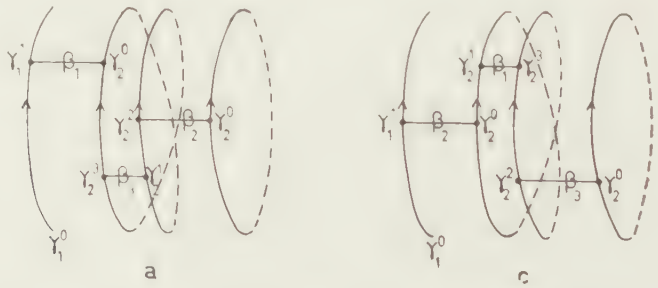

Fig. 6. Diagrams topologically equivalent to the diagrams of fig. 5 according to the arguments given under (3).

According to this the topological equivalence exists also between diagrams which contribute to different expansion coefficients $a_l^{(h)}(q'^h, q^h)$ of the reduced density matrix $n_h(q'^h, q^h)$.

Thus topologically equivalent diagrams can be transformed into each other by:

(1) interchange of the last l particle lines.

(2) a continuous displacement of the interaction end-points along the toron lines provided that no end-points cross on the same toron line.

(3) omitting or inserting additional windings in a toron line.

The consequence of the topological equivalence mentioned under (1) is a factor $l!$. The consequences of (2) and (3) will discussed hereunder. The main point is the discussion of (2) after which (3) is obtained without difficulties.

B) *Elimination of the non-topological features* (2)

We remark that the contribution $C(\Gamma^c_{h,h+l})$ of the diagrams calculated according to (23) is very similar for all diagrams of figs. 5a, b, c, d, 6a and c. The only differences are the factors of the type $\langle p_j{}^s|\mathscr{W}_1{}^0(\beta)|p_j{}^s\rangle$ in the integrand of (20) and the integration interval for the β_s.

In order to totalize over all diagrams which are topologically equivalent according to (2), it is convenient to consider the two end-points of an interaction line as having independent inverse temperatures β_s. Later on conditions will be imposed on these pairs of values of β_s to select just the correct diagrams.

In order to introduce these inverse temperatures for each of the interaction end-points, we number the interaction end-points with a pair of numbers (i, j), where j refers to the toron line of the interaction end-point ($1 \leqslant j \leqslant h$ for the reference lines and $j > h$ for the closed toron lines) and where i refers to the position of the interaction end-point on the toron line. The numbering of the positions i on the toron line is in the direction of increasing β. On the closed toron line the starting point $(0, j)$ may be chosen arbitrarily; on the reference lines we give for convenience the starting point of the reference line the number $(0, j)$ (although this is not an interaction end-point) and the first interaction end-point the number $(1, j)$. (i, j) numbers also the toron line segment pointing towards the interaction end-point (i, j).

Then we define for each interaction end-point (i, j) the *toron coordinate* $\gamma_j{}^i$, which is equal to the corresponding β_s of the interaction, except for a multiple of β. These multiples of β are chosen such that $\gamma_j{}^{i+1} - \gamma_j{}^i$ measures the "length" of the toron line segment, thus including the contribution β for each winding. For definiteness the value of $\gamma_j{}^0$ on the closed toron lines ($j > h$) is chosen between β and 0. The startingpoint of the reference line $j(\leqslant h)$ gets also a toron coordinate $\gamma_j{}^0 = 0$, which is convenient in treating the open and closed toron lines simultaneously.

The toron coordinates are already indicated in figs. 5 and 6.

Finally m_j is the number of complete windings in the toron line j.

Not all values of the γ_j^i correspond to real diagrams; in fact the following conditions have to be satisfied: the first set ensures the equality of the toron coordinates on both sides of the interaction line except for a multiple of β:

$$(36\alpha) \qquad \gamma_j{}^i = \gamma_{j'}{}^{i'} \pmod{\beta}$$

where $\gamma_j{}^i$ and $\gamma_{j'}{}^{i'}$ are the toron coordinates on both sides of the same interaction line. The other conditions account for the proper order of the toron coordinates on each toron line:

$$(36\beta) \qquad \gamma_j{}^0 + m_j\beta > \gamma_j{}^{l_j} > \ldots \gamma_j{}^1 > \gamma_j{}^0.$$

$\gamma_j{}^{l_j}$ is the last toron coordinate if we deal with a reference line and it is the one preceding $\gamma_j{}^0$ in the case of a closed toron line with m_j windings. We point out that (36_j) refers both to open ($\gamma_j{}^0 = 0$) and closed toron lines. The effect of these conditions will be indicated for the diagrams of figs. 5 and 6.

The condition (36β) for the diagrams of fig. 5 reads:

$$\gamma_1{}^0+\beta>\gamma_1{}^1>\gamma_1{}^0,\ \gamma_2{}^0+3\beta>\gamma_2{}^3>\gamma_2{}^2>\gamma_2{}^1>\gamma_2{}^0 \text{ and } \gamma_3{}^0+\beta>\gamma_3{}^0$$

where $\gamma_1{}^0 = 0$ and the multiples of β on the left hand side correspond to the number of windings in the toron lines, being 1, 3 and 1 respectively.

The conditions (36α) are satisfied for the various diagrams in the following way:

figs. 5a and b	$\gamma_1{}^1=\gamma_2{}^0$	$\gamma_2{}^2-2\beta=\gamma_3{}^0$	$\gamma_2{}^3-3\beta=\gamma_2{}^1-2\beta$
fig. 5c	$\gamma_1{}^1=\gamma_2{}^0$	$\gamma_2{}^2-2\beta=\gamma_3{}^0$	$\gamma_2{}^3-2\beta=\gamma_2{}^1-\beta$
fig. 5d	$\gamma_1{}^1=\gamma_2{}^0$	$\gamma_2{}^2-2\beta=\gamma_3{}^0$	$\gamma_2{}^3-3\beta=\gamma_2{}^1-\beta$.

As said $\gamma_j{}^0 (j=2, 3)$ is taken between 0 and β.

In the same way the conditions (36_j) read for the diagrams of fig. 6

$$\gamma_1{}^0+\beta>\gamma_1{}^1>\gamma_1{}^0,\ \gamma_2{}^0+2\beta>\gamma_2{}^3>\gamma_2{}^2>\gamma_2{}^1>\gamma_2{}^0 \text{ and } \gamma_3{}^0+\beta>\gamma_3{}^0$$

where again $\gamma_1{}^0 = 0$ and the multiples of β correspond to the number of windings in the toron lines, i.e. 1, 2 and 1 respectively.

The conditions (36α) read for the diagrams of fig. 6:

$$\text{figs. 6a and c} \qquad \gamma_1{}^1=\gamma_2{}^0 \qquad \gamma_2{}^2-\beta=\gamma_3{}^0 \qquad \gamma_2{}^3-\beta=\gamma_2{}^1.$$

The integration over the old inverse temperature coordinates β_s will be replaced by an integration over the new coordinates $\gamma_j{}^i$. The total number of $\gamma_j{}^i$ (aside from $\gamma_j{}^0 = 0$ for $j = h$) is just twice the number of β_s. The conditions (36α) reduce the number of independent variables again to half.

In calculating the contribution of the diagrams with the aid of the

toron coordinates we write the factor (i′) for the particle line segments between the interaction end-points (i, j) and $(i-1, j)$ (and with the quantum number labeled $p_j{}^i$) as:

$$f'_{\text{segm}} = \langle p_j{}^i | \mathscr{W}_1{}^0(\gamma_j{}^i - \gamma_j{}^{i-1}) | p_j{}^i \rangle \tag{i′}$$

refering now to the *toron line* segment (i, j). Of course at the ends of the reference lines the $p_j{}^i$ have to be replaced by k_j and $k'_{\text{P}(j)}$. We stress that (i′) includes the factors $\langle p_j{}^i | \mathscr{W}_1{}^0(\beta) | p_j{}^i \rangle$ due to the extra windings. The factor (ii′) remains:

$$f'_{\text{int}} = \langle p_j{}^i p_{j'}{}^{i'} | -\phi | p_j{}^{i-1} p_{j'}{}^{i'-1} \rangle. \tag{ii′}$$

Then the total contribution of all connected $(h, h+l)$ diagrams, contributing to the expression (25) for the coefficients $a_l{}^{(h)}(q'^h, q^h)$ in the expansion for the reduced density matrix, which are topologically equivalent because of the arguments given under (2), reads:

$$\int' d\gamma_1{}^1 \ldots \sum_{\{p_j{}^i\}} (\varepsilon^{h+l-c_{\text{P}}} \prod f'_{\text{segm}} f'_{\text{int}}) \tag{37}$$

where the prime on the integration means that the values of the $\gamma_j{}^i$ are subjected to the conditions (36α) and (36β), that $\gamma_j{}^0 = 0$ for $j \leqslant h$ and that $\beta > \gamma_j{}^0 > 0$ for $j > h$. The product now runs over the interaction lines and the toron line segments (instead of the particle line segments). The proof of (37) is based on the 1 : 1 correspondence between diagrams and sets of values of $\gamma_j{}^i$ satisfying these conditions (36α) and (36β). Indeed, a diagram can be built up unambiguously with the $\gamma_j{}^i$ starting with the reference lines and on each toron line from $\gamma_j{}^0$ (with $\beta > \gamma_j{}^0 > 0$). On the other hand the values of the toron coordinates $\gamma_j{}^i$ are defined uniquely in a given diagram again by the restriction $\beta > \gamma_j{}^0 > 0$ on the closed toron lines.

In order to facilitate the integration over the $\gamma_j{}^i$, the conditions (36α) and (36β) will be taken into account by introducing appropriate factors in the integrands of the contribution (37) as follows: the conditions (36α) are satisfied by introducing the δ-functions in the integrand for the $\gamma_j{}^i$ integration:

$$\delta(\gamma_j{}^i - \gamma_{j'}{}^{i'} \,(\text{mod}\, \beta)) = (1/\beta) \sum_{t_u = -\infty}^{\infty} \exp\{2\pi i t_u (\gamma_j{}^i - \gamma_{j'}{}^{i'})/\beta\} \tag{38α}$$

where the t_u are integers. The conditions (36β) are taken into account by writing, instead of the factor (i′), the integral expression:

$$\langle p_j{}^i | \mathscr{W}_1{}^0(\gamma_j{}^i - \gamma_j{}^{i-1}) | p_j{}^i \rangle = \int_{-\infty}^{\infty} ds_j{}^i \, \frac{\exp\{2\pi i s_j{}^i (\gamma_j{}^i - \gamma_j{}^{i-1})/\beta\}}{2\pi i s_j{}^i + \beta E(p_j{}^i)} \tag{38β}$$

where the integration over the $s_j{}^i$ runs along a path which goes below the zeroes of the denominator, i.e. on the lower side of the real axis. This equation (38β) in fact is only valid for $\gamma_j{}^i > \gamma_j{}^{i-1}$. As, however, the

right hand side is equal to zero for $\gamma_j{}^i < \gamma_j{}^{i-1}$, the introduction of this integral expression, which is due to MONTROLL and WARD [3] allows the conditions (36β) on the integration over $\gamma_j{}^i$ to be dropped. (For the last segment of a reference line we take the variable $s_j{}^0$).

The right hand side of (38β) is discontinuous and not defined at $\gamma_j{}^i = \gamma_j{}^{i-1}$. For the integration over $\gamma_j{}^i$ this is irrelevant, unless a factor $\delta(\gamma_j{}^i - \gamma_j{}^{i-1} \pmod\beta)$ occurs, as in the special case of a "self"-interaction in a toron line (i.e. when two successive end-points $\gamma_j{}^i$ and $\gamma_j{}^{i-1}$ belong to the same toron line). Such "self"-interactions may not contribute for the values $\gamma_j{}^i = \gamma_j{}^{i+1}$ as can be seen from the diagrams. The expressions can be made unambiguously in this case with the factor:

$$\exp(-2\pi i \eta s_j{}^i) \qquad \text{with } \eta \to 0^+ \tag{39}$$

in the integrand of (38β) which just excludes a contribution of the values $\gamma_j{}^i = \gamma_j{}^{i-1}$. This factor will be omitted in general for reasons of brevity.

The introduction of the expressions (38α) and (38γ) has the advantage that the integration over the $\gamma_j{}^i$ becomes elementary: since the restrictions (36α) and (36β) on the integration are now automatically satisfied, the

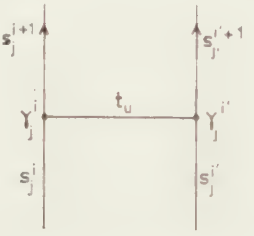

Fig. 7. Interaction line with toron coordinates $\gamma_j{}^i$ and $\gamma_{j'}{}^{i'}$. (cf. (40) and (40')).

integration over the $\gamma_j{}^i$ may be extended from $-\infty$ to $+\infty$, excepting the integration over $\gamma_j{}^0$ ($j > h$), which runs between 0 and β (and $\gamma_j{}^0 = 0$ for $j \leqq h$). We may consider one particular interaction u (see fig. 7) and collect all terms involving $\gamma_j{}^i$ and $\gamma_{j'}{}^{i'}$ at both end-points of the interaction ($i \neq 0, i' \neq 0$). The integration over $\gamma_j{}^i$ leads to:

$$\int_{-\infty}^{\infty} d\gamma_j{}^i \exp[2\pi i \gamma_j{}^i (s_j{}^i - s_j{}^{i+1} + t_u)/\beta] = \beta\delta(s_j{}^i - s_j{}^{i+1} + t_u) \tag{40}$$

and similarly the $\gamma_{j'}{}^{i'}$ integration gives:

$$\int_{-\infty}^{\infty} d\gamma_{j'}{}^{i'} \exp[2\pi i \gamma_{j'}{}^{i'} (s_{j'}{}^{i'} - s_{j'}{}^{i'+1} - t_u)/\beta] = \beta\delta(s_{j'}{}^{i'} - s_{j'}{}^{i'+1} - t_u). \tag{40'}$$

It should be stressed that the argument of the exponential is purely imaginary since $s_j{}^i$ and $s_j{}^{i-1}$ have the same imaginary part (c.f. (38β)). So (40) and (40') are proper expressions for the (Dirac) δ-functions. Thus each integration over a $\gamma_j{}^i$ ($i \neq 0$) gives a "conservation" law of the type:

$$s_j{}^i - s_j{}^{i+1} + t_u = 0. \tag{41}$$

The integration over $\gamma_j{}^0$ (on a *closed* toron line) is of the form:

$$(40^0) \qquad \int_0^\beta d\gamma_j{}^0 \exp\left[2\pi i\gamma_j{}^0(s_j{}^0 - s_j{}^1 + t_u)/\beta\right] = \beta\delta(s_j{}^1 - s_j{}^0, t_u)$$

because of the fact that by virtue of the conservation laws (41) in the toron line j

$$s_j{}^0 - s_j{}^1 = (s_j{}^0 - s_j{}^{l_j}) + \ldots (s_j{}^2 - s_j{}^1)$$

is equal to a sum of whole numbers t_u. The asymmetry due to the occurrence of one Kronecker δ (40^0) in a closed toron line among the Dirac δ-functions (40) and (40′) will be removed at the end of this section (E).

The integration over the toron coordinates $\gamma_j{}^0$ being performed, we summarize what is left from the expression (37) for the set of $(h, h+l)$ diagrams which are topologically equivalent because of the arguments mentioned under (2): the integrations over the $\gamma_j{}^0$ could be carried out at the cost of the introduction of the new integration variables $s_j{}^i$ associated with the toron line segments and of the summation variables t_u associated with the interaction lines. Then for each of the toron line segments we are left with the energy denominators of (38β):

$$f''_{\text{segm}} = 1/[2\pi i s_j{}^i + \beta E(p_j{}^i)]. \qquad \text{(i'')}$$

The factors β occurring in (38α) and (40) can be taken into account conveniently by including a factor β in each matrix element (ii′)

$$f''_{\text{int}} = \langle p_j{}^i p_{j'}{}^{i'} | -\beta\phi | p_j{}^{i-1} p_{j'}{}^{i'-1} \rangle. \qquad \text{(ii'')}$$

This may be seen to be correct on dimensional grounds only.

The integration over $s_j{}^i$ and the summation over t_u are restricted by δ-functions

$$f''_{\text{vertex}} = \delta(s_j{}^i - s_j{}^{i-1} \pm t_u) \qquad \text{(iii'')}$$

associated with each of the vertices (at each interaction *one* + and *one* − sign). For each *closed* toron line one of these δ-functions, which can be chosen arbitrarily, must be replaced by a Kronecker δ.

Then the expression (37) for the total contribution of a set of diagrams, which are topologically equivalent due to (2), becomes:

$$(42) \qquad \int ds_1{}^0 \ldots \sum_{\{p_j{}^i\}} \sum_{\{t_u\}} (\varepsilon^{h+l-c_P} \prod f''_{\text{segm}} f''_{\text{int}} f''_{\text{vertex}} \exp(2\pi i s_j{}^0 m_j)).$$

We remark that the factors $\exp(2\pi i s_j{}^0 m_j)$ in (42) which are due to the first condition of (36β), are the only ones which depend on the number of windings m_j in each of the toron lines j.

C) *Elimination of the non-topological features* (3)

Now we discuss the topological equivalence as mentioned under (3), which is concerned with topologically equivalent diagrams with different sets of numbers of windings m_j. These sets of numbers m_j are subjected

to the restriction $\sum_j m_j = h+l$, when we take only the $(h, h+l)$ diagrams which contribute to the expression (25) for the l-th coefficient $a_l^{(h)}(q'^h, q^h)$ in the expansion of the reduced density matrix $n_h(q'^h, q^h)$. Therefore it is easier to calculate directly the reduced density matrix itself instead of $a_l^{(h)}(q'^h, q^h)$, because then the summation over l removes the restriction $\sum_j m_j = h+l$.

So we collect the factors depending on l and m_j in (42) and multiply with the $(h+l)$-th power of $\bar{z}$ according to the expansion (27) of the reduced density matrix: They may be written in the form:

$$(43)\quad \varepsilon^{h+l-c_P}\Big(\prod_{j=1}^{g}(\exp 2\pi i s_j^0)^{m_j}\Big)\,\bar{z}^{h+l} = \varepsilon^{g-c_P}\prod_{j=1}^{g}\varepsilon^{-1}[\varepsilon\exp(\beta\mu + 2\pi i s_j^0)]^{m_j}$$

where g is the total number of (open and closed) toron lines. In the unrestricted summation over m_j each m_j runs from 1 to ∞. Therefore we obtain from (43)

$$\varepsilon^{g-c_P}\prod_{j=1}^{g}\frac{\exp(\beta\mu+2\pi i s_j^0)}{1-\varepsilon\exp(\beta\mu+2\pi i s_j^0)}$$

i.e. for each toron line the factor:

$$f''_{\text{toron}} = \frac{\exp(\beta\mu+2\pi i s_j^0)}{1-\varepsilon\exp(\beta\mu+2\pi i s_j^0)} \qquad \text{(iv'')}$$

provided that $|\varepsilon\exp\beta\mu|<1$. We remark that $f''_{\text{toron}}(s_j^0)$ is periodic in s_j^0 with the period 1, such that every s_j^i of the j-th toron line may be inserted in (iv'') since these s_j^i of the j-th toron line differ only whole numbers due to the conservation laws (41). This completes the calculation of the contribution of the sum of all topologically equivalent diagrams.

D) *Introduction of the primary h-diagrams*

In order to give a representation of the results, we define a new type of diagrams: the so called "*primary h-diagrams*" which are determined by their topological properties only. Such a primary diagram then should represent the sum of the contribution of the corresponding set of topologically equivalent diagrams. The name 'primary' is taken over from Yang and Lee [6] as the primary diagrams are at the beginning of the same type of developments as the "primary graphs" of Yang and Lee.

The primary h-diagrams consist of h open toron lines (reference lines) and a number of closed toron lines connected by interaction lines. At the beginnings of the h reference lines are the numbers $1 \ldots h$ and at the ends a permutation $P(1) \ldots P(h)$. The toron line segments have an arrow and are labeled by (i, j), the numbers i and j numbering both the interaction end-points and the toron line segment pointing towards the vertex (i, j). To each toron line segment (i, j) are associated the quantum numbers p_j^i and the variables s_j^i. The begin segments of the reference lines have the quantum numbers $k_1 \ldots k_h$ and the variables $s_1^1 \ldots s_h^1$,

whereas the end segments have the quantum numbers $k_1 \ldots k_h$ and the variables $s_1{}^0 \ldots s_h{}^0$. The interaction lines are drawn as broken lines and a variable t_u, which takes only integer values, is associated to each of these interaction lines. Then we have:

$$n_h(q'^h, q^h) = \sum_{k'^h} \sum_{k^h} \langle q'^h | k'^h \rangle \left[\sum_{\Gamma_h{}^p} C''(\Gamma_h{}^p) \right] \langle k^h | q^h \rangle \tag{44}$$

where the contribution $C''(\Gamma_h{}^p)$ of a primary h-diagram $\Gamma_h{}^p$ is calculated as:

$$C''(\Gamma_h{}^p) = \int ds_1{}^0 \ldots \sum_{\{p_j{}^i\}} \sum_{\{t_u\}} (\varepsilon^{h - c_P(h)} \prod f''_{\text{segm}} f''_{\text{int}} f''_{\text{vertex}} f''_{\text{toron}})_{\Gamma_h{}^p} \tag{45}$$

$c_P(h)$ is the number of cycles in the permutation P(1), ... P(h). (We note that $h - c_P(h)$ is equal to $g - c_P$ (c.f. (43)) as every *closed* toron line corresponds to a cycle in the permutation P.) To illustrate (45) we have

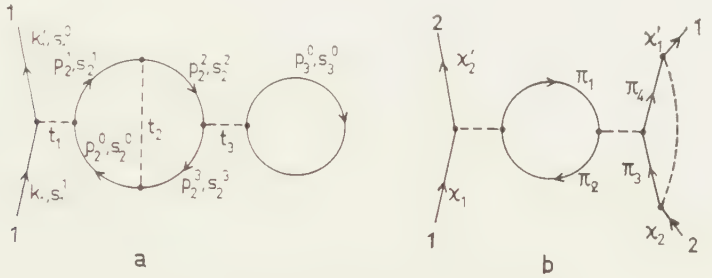

Fig. 8. Topologically invariant diagrams (c.f. (46) and (63)).

drawn in fig. 8a the topological representant of the diagrams of figs. 5 and 6, and in fig. 8b that of the diagram fig. 3b. The contribution of the diagram fig. 8a reads:

$$\begin{aligned} &\Bigg\{ \int ds_1{}^0 \ldots \sum_{\{p_j{}^i\}} \sum_{\{t_u\}} \frac{1}{2\pi i s_1{}^1 - \beta E(k_1)} \frac{1}{2\pi i s_1{}^0 - \beta E(k_1')} \langle k_1' p_2{}^1 | \quad \beta\phi \, k_1 p_2{}^0 \\ &\Bigg\{ \prod_{i=0}^{3} \frac{1}{2\pi i s_2{}^i - \beta E(p_2{}^i)} \Bigg\} \langle p_2{}^2 p_2{}^0 \cdot \beta\phi \; p_2{}^1 p_2{}^3 \cdot p_2{}^3 p_3{}^0 \cdot \quad \beta\phi \; p_2{}^2 p_3{}^0 \\ &\Bigg\{ \frac{1}{2\pi i s_3{}^0 + \beta E(p_3{}^0)} \, \delta(s_1{}^0 - s_1{}^1 - t_1) \, \delta(s_2{}^0 - s_2{}^1, t_1) \, \delta(s_2{}^1 - s_2{}^2 + t_2) \\ &\delta(s_2{}^2 \quad s_2{}^3 - t_3) \, \delta(s_2{}^3 \quad s_2{}^0 - t_2) \, \delta(s_3{}^0 \quad s_3{}^0, t_3) \, f''_{\text{toron}}(s_1{}^0) \, f''_{\text{toron}}(s_2{}^0) \, f''_{\text{toron}}(s_3{}^0). \end{aligned} \tag{46}$$

E) *The integration over $s_j{}^i$*

The new integration variables $s_j{}^i$ show some similarity to the quantum numbers $p_j{}^i$ containing the momentum of the particles. We will show now that the summation over t_u and the integration over $s_j{}^i$ can be combined with the summation over the quantum numbers $p_j{}^i$ by extending the momentum variables with a fourth component. Therefore we investigate the integration over the variables $s_j{}^i$.

In each toron line j one can integrate over $s_j{}^1, s_j{}^2, \ldots$ (except $s_j{}^0$), by

using the (Dirac) δ-functions of the vertices. The result is that in the expressions for f''_{segm} in (45) $s_j{}^1, s_j{}^2 \ldots$ are expressed as functions of $s_j{}^0$ and the numbers t_u of the interaction lines ending on the j-th toron line (see fig. 9). In a closed toron line then the (Kronecker) δ of the last vertex leads to an expression between the t_u of the interaction lines only. So the integration over $s_j{}^0$, which thus now occurs in each f''_{segm} of the j-th toron line and in f''_{toron}, is not restricted by a δ-function (both for the closed and open toron lines).

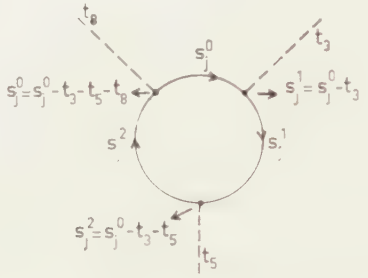

Fig. 9. Expressions for the quantities $s_j{}^i$ in terms of $s_j{}^0$ and t_u with the aid of the δ-functions of the vertices.

In order to carry out the integration over $s_j{}^0$, we have to consider the analytical behaviour of the functions f''_{segm} and f''_{toron}. The poles of

$$f''_{\text{segm}}(s_j{}^i) = \frac{1}{2\pi i s_j{}^i + \beta E(p_j{}^i)}$$

are given by:

$$2\pi i s_j{}^0 = 2\pi i (\sum t_u) - \beta E(p_j{}^i) \tag{47}$$

where $\sum t_u$ is the difference $s_j{}^i - s_j{}^0$, which results when $s_j{}^i$ is expressed in $s_j{}^0$ and the t_u. They lie in the *upper* half plane ($\beta E(p_j{}^i) > 0$) on lines with integer real parts (namely equal to $\sum t_u$). The poles of

$$f''_{\text{toron}}(s_j{}^0) = \frac{\exp(\beta\mu + 2\pi i s_j{}^0)}{1 - \varepsilon \exp(\beta\mu + 2\pi i s_j{}^0)}$$

are given by:

$$2\pi i s_j{}^0 = 2\pi i t_j{}^0 - \beta\mu - \ln\varepsilon \qquad t_j{}^0 = 0, \pm 1, \ldots . \tag{48}$$

They lie at equidistant points with an interval 1 on a line parallel to the real axis in the *lower* half plane (since we assumed $|\varepsilon \exp \beta\mu| < 1$ for the validity of the summation (43); the case $\mu = 0$ will be discussed at the end of this section). We stress that the poles of f''_{toron} for $\varepsilon = 1$ (B – E) have integer real parts, whereas for $\varepsilon = -1$ (F – D) the real parts of the poles of f''_{toron} are moved over a distance $\frac{1}{2}$, due to the imaginary part of $\ln \varepsilon$ (see figs. 10a and b). To treat the Boltzmann statistics on the same footing we must take ε finite and take afterwards the limit $\varepsilon \to 0$.

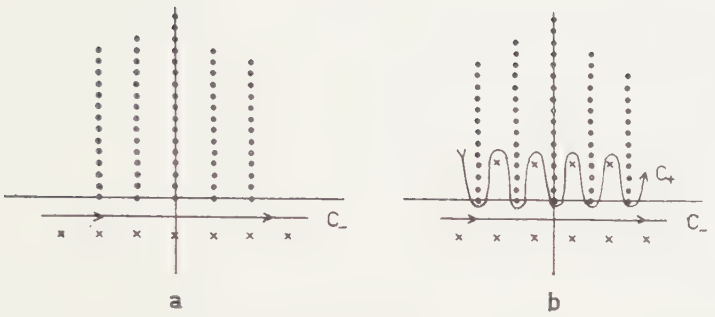

Fig. 10. Poles . and × of the functions f''_{segm} and f''_{toron} respectively in the case of Bose-Einstein statistics (10a) and Fermi-Dirac statistics (10b).

The integration path C for $s_j{}^0$ runs *below* the poles of the f''_{segm} factors and *above* the poles of f''_{toron}. Now the f''_{segm} functions behave like $1/2\pi i s_j{}^0$ at infinity and f''_{toron} $(s_j{}^0)$ decreases exponentially in the upper half plane and approaches to $-(1/\varepsilon)$ in the lower half plane. So it is always permitted to close the integration path with a large semi-circle through the *upper* half plane and the integration path may be closed through the *lower* half plane when *two or more* factors f''_{segm} are present. It is important to note that also in all cases where an open or closed toron line leads to only *one* factor f''_{segm}, the integration path can be closed through the lower half plane after introduction of the convergence factor

$$\exp -2\pi i\eta s_j{}^0 \qquad \text{with } \eta \to 0^+ \tag{49}$$

which is the same as the factor (39) which has to be added to the factor f''_{segm} corresponding to the interactionless toron line segments between the end-points of a "self"-interaction.

We illustrate the method of closing the integral over the upper or over the lower half plane with the example of the 1-diagram without interaction. The contribution of this diagram to $n_1(q_1', q_1)$ (denoted by $n_1{}^0(q_1', q_1)$) is given by:

$$n_1{}^0(q_1', q_1) = \sum_k \langle q_1'|k\rangle \left[\int ds \frac{1}{2\pi i s - \beta E(k)} \frac{\exp(\beta\mu + 2\pi i s)}{1 - \varepsilon \exp(\beta\mu - 2\pi i s)}\right] \langle k|q_1\rangle. \tag{50}$$

Closing the s-integration through the *upper* half plane gives:

$$n_1{}^0(q_1', q_1) = \sum_k \langle q_1'|k\rangle\, n^0(k)\, \langle k|q_1\rangle \tag{51}$$

with

$$n^0(k) = \frac{\exp \beta(\mu - E(k))}{1 - \varepsilon \exp \beta(\mu - E(k))} \tag{52}$$

$n^0(k)$ is the well-known occupation number density (in zeroth approximation) which applies to the three types of statistics.

Closing on the other hand the s integration through the *lower* half plane, we rewrite (50) including the convergence factor (49) and calculate the residues at the poles (48) of $f''_{\text{toron}}(s)$. In the neighbourhood of these poles $f''_{\text{toron}}(s)$ behaves as:

$$f''_{\text{toron}}(s) = \frac{-(1/\varepsilon)}{2\pi i s - (2\pi i t - \ln \varepsilon - \beta\mu)} + \ldots \qquad t = 0, \pm 1, \ldots \tag{53}$$

So we arrive at (51) with the expression for $n^0(k)$:

$$n^0(k) = \sum_{t=-\infty}^{\infty} (1/\varepsilon) \frac{\exp -\eta(2\pi i t - \ln \varepsilon - \beta\mu)}{2\pi i t - \beta\mu - \ln \varepsilon + \beta E(k)} \qquad \eta \to 0^+ \tag{54}$$

This fourier sum on the righthand side may be seen to give (52) by writing the denominator as:

$$[2\pi i t - \beta\mu - \ln \varepsilon - \beta E(k)]^{-1} = \int_0^1 d\alpha \, \frac{\exp -\alpha(2\pi i t - \beta\mu - \ln \varepsilon + \beta E(k))}{1 - \varepsilon \exp \beta(\mu - E(k))} \tag{55}$$

After inserting (55) into (54) and interchanging of summation and integration the summation over t gives the factor

$$\sum_{t=-\infty}^{\infty} \exp -2\pi i t(\alpha + \eta) = \delta(\alpha + \eta \ (\text{mod } 1)). \qquad \eta \to 0^+$$

For the integration over α this implies that we have to put $\alpha = 1 - \eta$ ($\eta \to 0^+$) yielding the same expression (52) for $n^0(k)$.

We will use the possibility of closing the s^0-integration path through the *lower* half plane, because then the rules for the calculation of the contribution of the diagrams becomes more symmetric. Using the expansion (53) of $f''_{\text{toron}}(s, s^0)$ in the neighbourhood of the poles (48), it can be seen that the residues of the product: $f''_{\text{toron}}(s, s^0) \prod f''_{\text{segm}}(s_j{}^t)$ in these poles are obtained by inserting $2\pi i s_j{}^t = 2\pi i t_j{}^t - \ln \varepsilon - \beta\mu$ into the factors $f''_{\text{segm}}(s_j{}^t)$ giving:

$$n^0(p_j{}^t, t_j{}^t) = 1/[2\pi i t_j{}^t - \ln \varepsilon - \beta\mu + \beta E(p_j{}^t)] \tag{56}$$

and multiplying the whole by $(1/\varepsilon)$ (due to f''_{toron}). The $t_j{}^t$ are obtained from $t_j{}^0$ and the t_u of the interaction lines ending on the j-th toron line in the same way as the $s_j{}^t$ from $s_j{}^0$ and the t_u (see fig. 9).

It is now convenient however to use as summation variables the $t_j{}^t$ instead of the t_u (and $t_j{}^0$). In order to express the t_u in terms of the $t_j{}^t$, we need one conservation law for each interaction line. Each interaction line has, however, two vertices and we are left with a relation between the incoming $t_j{}^t$ and $t_{j'}{}^{t'}$ on the one side and the outgoing $t_j{}^{t+1}$ and $t_{j'}{}^{t'+1}$ on the other side of the interaction. The two conservation laws $t_j{}^t - t_j{}^{t+1} = t_u$ and $t_{j'}{}^{t'} = t_{j'}{}^{t'+1} - t_u$ result in one between the $t_j{}^t$ which may be expressed by associating the factor

$$\delta(t_j{}^t + t_{j'}{}^{t'}, \ t_j{}^{t+1} + t_{j'}{}^{t'+1}) \tag{57}$$

to each interaction line. The conservation law for the t_j^i is analoguous to the conservation law for the total momentum, which is included (implicitely) in the matrix elements of the interaction. The variable t_j^i thus plays the role of an "energy" coordinate, which can be conveniently combined with the quantum numbers p_j^i. Thus the factor (57) can be included in the matrix element of the interaction by defining:

$$\begin{cases} \langle \pi_j^i \, \pi_{j'}^{i'} | -\beta\phi | \pi_j^{i+1} \, \pi_{j'}^{i'+1} \rangle = \\ \quad = \langle p_j^i \, p_{j'}^{i'} | -\beta\phi | p_j^{i+1} \, p_{j'}^{i'+1} \rangle \, \delta(t_j^i + t_{j'}^{i'}, \, t_j^{i+1} + t_{j'}^{i'+1}) \end{cases} \tag{58}$$

where we used the notation:

$$\pi_j^i = (p_j^i, \, t_j^i). \tag{59}$$

In summing up the final rules for the contribution of the primary h-diagrams we drop the double indices by indicating the segments, their quantum numbers and "energies". Each internal toron line segment carries a label j and an associated momentum-"energy" coordinate $\pi_j = (p_j, t_j)$. The external toron line segments of the reference lines carry the parameters $\varkappa'_{\mathrm{P}(1)}, \varkappa'_{\mathrm{P}(2)} \ldots \varkappa'_{\mathrm{P}(h)}$ for the outgoing lines and the parameters $\varkappa_1, \varkappa_2 \ldots \varkappa_h$ for the incoming lines.

To the toron line segment j corresponds the *particle propagator*:

$$n^0(\pi_j) = 1/[2\pi i t_j - \beta\mu - \ln \varepsilon + \beta E(p_j)]. \tag{56}$$ (i)

It should be remembered that in all cases where an open or closed toron line leads to only *one* particle propagator or when an interactionless toron line segment occurs between the end-points of the same self-interaction, a convergence factor (see (49) and (39)) has to be included

$$f_{\mathrm{conv}} = \exp(-2\pi i \eta t_j) \qquad \eta \to 0^+. \tag{i}$$

To the interaction lines corresponds the factor (58):

$$f_{\mathrm{int}} = \langle \pi_{j_1} \, \pi_{j_2} | -\beta\phi | \pi_{j_3} \, \pi_{j_4} \rangle. \tag{ii}$$

To each toron line as a whole corresponds a factor:

$$f_{\mathrm{toron}} = 1/\varepsilon. \tag{iii}$$

The reduced density matrices can now be expressed in the form:

$$n_h(q'^h, q^h) = \sum_{\varkappa'^h} \sum_{\varkappa^h} \langle q'^h | k'^h \rangle \, n_h(\varkappa'^h, \varkappa^h) \, \langle k^h | q^h \rangle \tag{60}$$

where the functions $n_h(\varkappa'^h, \varkappa^h)$ are defined as:

$$n_h(\varkappa'^h, \varkappa^h) = \sum_{\Gamma_h^{\mathrm{p}}} C(\Gamma_h^{\mathrm{p}}) \tag{61}$$

and $C(\Gamma_h^{\mathrm{p}})$ is the contribution of the primary h-diagram Γ_h^{p} calculated as

$$C(\Gamma_h^{\mathrm{p}}) = \prod_{m=1}^{h} n^0(\varkappa_m) \, n^0(\varkappa'_m) \sum_{\{\pi_j\}} (\varepsilon^{h - c_{\mathrm{P}}(h)} \prod n^0(\pi_j) \, f_{\mathrm{int}} \, f_{\mathrm{toron}})_{\Gamma_h^{\mathrm{p}}}. \tag{62}$$

To illustrate the new rules we give the contribution of the diagram of fig. 8b which reads:

$$(63)\quad \Big\} : n^0(\varkappa_1)\, n^0(\varkappa_2)\, n^0(\varkappa_1')\, n^0(\varkappa_2') \Big\{ \left(\frac{1}{\varepsilon}\right)^3 \sum_{\pi_1 \ldots \pi_4} \left(\prod_{j=1}^{4} n^0(\pi_j)\right) \langle \varkappa_2' \pi_1 | -\beta\phi | \varkappa_1 \pi_2 \rangle$$
$$\langle \pi_2 \pi_4 | -\beta\phi | \pi_1 \pi_3 \rangle \langle \varkappa_1' \pi_3 | -\beta\phi | \pi_4 \varkappa_2 \rangle .$$

From the functions $n_h(\varkappa'^h, \varkappa^h)$ the most important one is $n_1(\varkappa_1', \varkappa_1)$. This function contains the δ-factor $\delta(\varkappa_1', \varkappa_1)$ which is the combined result of all conservations of total momenta, spin and "energies" in the diagram. We write therefore:

$$(64)\qquad n_1(\varkappa_1', \varkappa_1) = (1/\varepsilon)\, \delta(\varkappa_1', \varkappa_1)\, n(\varkappa_1).$$

The function $n(\varkappa)$ will be used intensively in the following papers. $n(\varkappa)$ is related to the number of particles by:

$$(65)\qquad N = \sum_k n(k) = \frac{1}{\varepsilon} \sum_{kt} n(\varkappa) \exp(-2\pi i \eta t) \qquad \eta \to 0^+$$

where the convergence factor (ī) is used.

Remark 1. We have to consider separately the question what happens if μ approaches the value 0. As we know that $\mu \ll 0$ for high temperatures this question becomes of importance for low temperatures. We see from the expression (52) for $n^0(k)$ that for Fermi–Dirac statistics ($\varepsilon = -1$) nothing happens, but that $n^0(k)$ becomes singular for $E(k) = 0$ in the case of Bose–Einstein statistics ($\varepsilon = 1$).

With respect to the $s_j{}^0$-integration we see from fig. 10b that for $\varepsilon = -1$ (F–D) the poles of $f''_{\text{toron}}(s_j{}^0)$ come near to the real axis when $\mu \to 0$, but do not interfere with the poles of the segment functions f''_{segm}. We may deforme the integration path C to a path C' such that it still runs *above* the poles of f''_{toron} and below the poles of f''_{segm}. The contributions of the diagrams calculated in this way are the analytical continuation of the power series in $\bar{z} = \exp \beta\mu$ for $|\bar{z}| > 1$.

When μ approaches 0 in the case of $\varepsilon = 1$ (B–E) the integration path becomes locked in between the poles (47) of the f''_{segm} factors and the poles (48) of f''_{toron}, both having integer real parts. No analytical continuation is possible; the state with $E(k_j{}^t) = 0$ gives an infinite contribution. In (52) the occupation of the state $E(k) = 0$ becomes singular i.e. comparable with the number of particles in the system. This is the way in which the well-known Bose Einstein condensation phenomenon appears in this description. We postpone the treatment of the Bose Einstein system below the condensation point to a future paper.

Remark 2. We stress that due to the closure of the $s_j{}^0$-integration through the *lower* half plane the new rules automatically furnish the analytical continuation of the diagram contributions in the Fermi-Dirac case.

For Bose–Einstein statistics the particle propagator $n^0(\pi)$ becomes singular for the state $E(p)=0$, $t=0$ when μ approaches 0. In dealing with the degenerate Bose gas this state must be treated separately.

The new rules are not very appropriate for the Boltzmann statistics because of the occurrence of $(1/\varepsilon)$ and $\ln \varepsilon$. When, however, the summation over the "energy"-coordinates t_j in the contribution of a diagram is carried out, the Boltzmann results follow inmediately by putting $\varepsilon=0$.

§ 5. Primary 0-diagrams for the partition function

In this section we shall discuss the introduction of topological diagrams related to the $(0, l)$ diagrams, used in § 3 for the calculation of the grand canonical partition function Z_{gr}. We do not follow the arguments given for the h-diagrams to obtain primary 0-diagrams. On the contrary, we define the primary 0-diagrams and prove that their contributions amount to $\ln Z_{\text{gr}}$.

A primary 0-diagram consists of closed toron lines (with arrow) connected by (broken) interaction lines. The toron line segments carry a label j and the parameter $\pi_j(=p_j, t_j)$. First we consider the diagrams with at least *one* interaction line, as the single toron line without interactions needs a special treatment. The following factors are associated to the diagrams with interaction:

corresponding to the toron line segment j the particle propagator:

$$n^0(\pi_j) = 1/[2\pi i t_j - \beta\mu - \ln \varepsilon + \beta E(p_j)] \tag{i}$$

completed when necessary (c.f. (39) and (49)) with the convergence factor

$$f_{\text{conv}} = \exp(-2\pi i \eta t_j) \tag{ī}$$

corresponding to the interaction lines the matrix element:

$$f_{\text{int}} = \langle \pi_{j_1} \pi_{j_2} | -\beta\phi | \pi_{j_3} \pi_{j_4} \rangle \tag{ii}$$

corresponding to the toron lines the factor

$$f_{\text{toron}} = 1/\varepsilon \tag{iii}$$

the diagram as a whole gets the symmetry factor

$$f_{\text{symm}} = 1/S \tag{iv}$$

where the symmetry number S is defined as follows: consider the labels j of the toron line segments. The number of permutations of the labels that leave the diagram topologically invariant (including the relative position of these labels) is the symmetry number S. In this definition the diagram is defined to be topologically invariant with respect to label permutations if the diagram with the permutated labels can be transformed into the original one by continuous deformation (the direction of the arrows is of importance!). We note that this definition for the

symmetry number S gives $S = 1$ for all primary h-diagrams, since due to the reference line(s), the arrows in the toron line segments and the connectedness of the diagram, each toron line segment has an unique position in the primary h-diagram. In fig. 11 we give some primary 0-diagrams and their symmetry number S.

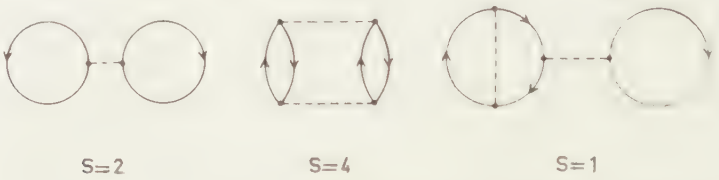

Fig. 11. Primary 0-diagrams with their symmetry numbers S.

With these definitions it will be proved that

$$\beta pV = \beta p^0 V + \sum_{\Gamma_0^{\mathrm{p}}} C(\Gamma_0^{\mathrm{p}}) \tag{66}$$

where the contribution $C(\Gamma_0^{\mathrm{p}})$ of the primary 0-diagram Γ_0^{p} is calculated as:

$$C(\Gamma_0^{\mathrm{p}}) = \sum_{\{\pi_j\}}{}' \left(\prod n^0(\pi_j)\, f_{\mathrm{int}}\, f_{\mathrm{toron}}\, f_{\mathrm{symm}}\right)_{\Gamma_0^{\mathrm{p}}}. \tag{67}$$

${}'\beta p^0 V$ is the ideal gas contribution to the pressure and the prime on the summation sign reminds of the fact that the diagram consisting of the single toron line without interaction is not included.

Before proving (66) we give the contribution of the ideal gas to the expression (29) for ${}'\beta pV$, coming from all connected $(0, l)$ diagrams without interactions. The summation over all these diagrams (see e.g. Bloch and de Dominicis [2]) leads to the expression:

$$\beta p^0 V = -(1/\varepsilon) \sum_k \ln\left[1 - \varepsilon \exp \beta(\mu - E(k))\right]. \tag{68}$$

For the proof of the expression (66) for βpV, we remark that the sum over the diagrams Γ_0^{p} in the right hand side of (66) is a functional of $n^0(\pi)$, which will be denoted by:

$$\beta V \boldsymbol{p}_l(n^0(\pi)) = \sum_{\Gamma_0^{\mathrm{p}}} \sum_{\{\pi_j\}}{}' \left(f_{\mathrm{symm}} \prod n^0(\pi_j)\, f_{\mathrm{int}}\, f_{\mathrm{toron}}\right)_{\Gamma_0^{\mathrm{p}}}. \tag{69}$$

First the following lemma is proved: concerning $n(\pi)$ defined in (64) and the functional derivative of $\boldsymbol{p}_l$ with respect to $n^0(\pi)$:

$$\text{lemma} \quad (1/\varepsilon)\left[n(\varkappa) - n^0(\varkappa)\right] = \left[n^0(\varkappa)\right]^2 \beta V \frac{\delta\, \boldsymbol{p}_l(n^0(\pi))}{\delta\, n^0(\varkappa)}. \tag{70}$$

The proof of this lemma is based on an important relationship between primary 0-diagrams contributing to the righthand side and primary 1-diagrams contributing to the lefthand side. If we take a primary 0-dia-

gram we obtain primary 1-diagrams by cutting one of the toron line segments and taking the cut toron line as reference line. To every 0-diagram a class of related 1-diagrams exists, arising from a cut in each of the toron line segments j (see figs. 12a and b). Two different 0-diagrams give different classes of related 1-diagrams. In a class of 1-diagrams which is related to one particular 0-diagram with symmetry number S there are S identical 1-diagrams. The reason for this is the fact that each permutation of the labels leaving the diagram topologically invariant changes *all* labels, because a toron line segment with an unreplaced label would work as a reference line in the diagram and reduce such a permutation to the identical permutation. So the S permutations, leaving the diagram invariant, divide the toron line segments into classes of S equivalent toron line segments; each of the equivalent segments giving the same 1-diagram when the toron line segments are cut open.

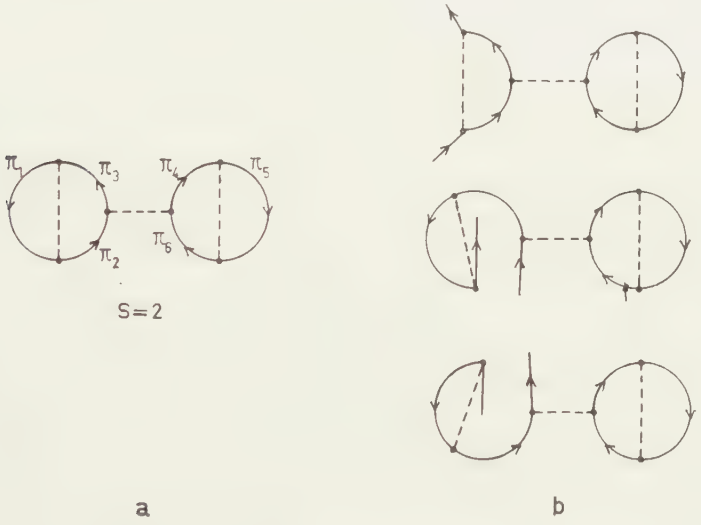

Fig. 12. Primary 0-diagrams and the related primary 1-diagrams (c.f. (72)).

Now taking the functional derivative of $\boldsymbol{p}_i$ with respect to $n^0(\pi)$ means that successively a toron line segment is singled out, carrying a factor $n^0(\pi)$ in the expression (70) for $\boldsymbol{p}_i$. So the relationship between 0-diagrams and 1-diagrams can be used and we see that all 1-diagrams are produced, while the symmetry factor $(1/S)$ given by (iv) takes care of the correct number of 1-diagrams. The factor $[n^0(\pi)]^2$ in the right hand side of the lemma furnishes the particle propagators for the external toron line segments of the 1-diagrams.

With the use of the lemma (70) equation (66) is easily proved by taking the derivatives of both sides in respect of μ. The derivative of $\boldsymbol{p}_i$ can be calculated with the aid of the lemma since μ occurs only through the factor $n^0(\pi)$. So:

$$\frac{\partial}{\partial\mu}\beta V\mathfrak{p}_i(n^0(\pi)) = \beta V\sum_{\varkappa}\frac{\partial\mathfrak{p}_i}{\partial n^0(\varkappa)}\frac{\partial n^0(\varkappa)}{\partial\mu} = \tag{71}$$

$$\beta V\sum_{\varkappa}\frac{\partial\mathfrak{p}_i}{\partial n(\varkappa)}\Big|_{n^0(\varkappa)}\left(\frac{\beta}{\varepsilon}\right)\sum_{\varkappa}(n(\varkappa)-n^0(\varkappa)).$$

Thus taking the derivative of both sides of (66) gives with (68) and (71)

$$\frac{\partial\beta pV}{\partial\mu} = \beta\sum_{k}n^0(k)+(\beta/\varepsilon)\sum_{\varkappa}(n(\varkappa)-n^0(\varkappa))\exp(-2\pi i\eta t) = \beta N \qquad (\eta\to 0^+).$$

This equality is identical with equation (33). Therefore both sides of (66) are equal to each other up to a constant independent of μ. This constant is zero as can be seen by taking the limit $\mu\to-\infty$. As the interaction effects vanish for $\mu\to-\infty$ we are than left with the interactionless case for which (66) is obviously true.

As an example of the rules (i)–(iv) we give the contribution of the diagram drawn in fig. 12a, reading:

$$\frac{1}{2!}\left(\frac{1}{\varepsilon}\right)^2\sum_{\pi_1\ldots\pi_6}\left(\prod_{j=1}^{6}n^0(\pi_j)\right)\langle\pi_1\pi_2|-\beta\phi|\pi_3\pi_1\rangle\langle\pi_3\pi_1|-\beta\phi|\pi_2\pi_6\rangle \tag{72}$$

$$\langle\pi_6\pi_5|-\beta\phi|\pi_5\pi_4\rangle.$$

Discussion

Topologically invariant diagrams are obtained for the reduced density matrices and the partition function. As regards the latter these primary 0-diagrams are identical to the diagrams introduced by Montroll and Ward [3] (c.f. also Luttinger and Ward [7]). The name primary is taken over from Lee and Yang because in subsequent papers further developments will be made analoguous to Lee and Yang's developments with the "primary graphs". (Our diagrams come most close to their "*contracted* graphs"; the difference between "contracted" and "primary graphs" is, however, small.)

The diagrams of Montroll and Ward as represented in this section have similar properties as Yang and Lee's "primary graphs", but there are noteworthy differences in the set up of the both types of diagrams. Yang and Lee divide the many body problem into two parts: one concerning the interaction which may be called the Boltzmann problem and one concerning the symmetry effects.

The Boltzmann problem amounts to the calculation of the matrix elements of the $\mathscr{U}_l$ operators resulting from the Ursell expansion of $\exp -\beta\mathscr{H}_N$. This is done with the (non-topological) diagrams, which formed the starting point of this paper (§ 2). Then the symmetry effects are treated with the so called "primary graphs", containing the interactions through the matrix elements of the Boltzmann operators. These "primary graphs" are topological diagrams, but their interaction points represent complicated functions calculated by non-topological diagrams.

In our approach the Boltzmann case is not taken as an essential inter-

mediate stage. The interaction effects appear in the diagrams through simple functions (matrix elements of the interaction), and the symmetry effects demonstrate themselves in the topological structure of the diagrams.

To our opinion there is not much advantage from a *mathematical* point of view to give the Boltzmann case an essential place, because the calculation of the Boltzmann $\mathscr{U}_l$ matrix elements involves the same type of integrals as the calculation of the "primary graph" contributions.

From a *physical* point of view it is advantageous to take the Boltzmann case as starting point when the symmetry effects are small. We have the feeling, however, that this advantage becomes doubtful when the symmetry effects are important. It turns out that, following YANG and LEE's method of treating the degenerate Bose gas, but using MONTROLL and WARD's diagrams as given here, one comes to expressions which can be interpreted more directly in terms of the collective excitations (phonons) of the degenerate Bose gas.

The author is much indebted to Prof. Dr. J. DE BOER for his inestimable advice and a revision of the manuscript.

REFERENCES

1. LEE, T. D. and C. N. YANG, Phys. Rev. **113**, 1165 (1959).
   ————, Phys. Rev. **116**, 25 (1959).
   —, Phys. Rev. **117**, 12 (1960).
2. BLOCH, C. and C. DE DOMINICIS, Nucl. Phys. **10**, 181 (1959).
   ————, Nucl. Phys. **7**, 459 (1958).
3. MONTROLL, E. W. and J. C. WARD, Phys. of Fuids **1**, 55 (1958).
4. SALPETER, E. E., Ann. of Phys. **5**, 183 (1958).
5. BOER, J. DE, Rep. Prog. Phys. **13**, 305 (1948/49).
6. LEE, T. D. and C. N. YANG, Phys. Rev. **117**, 22 (1960).
7. LUTTINGER, J. M. and J. C. WARD, Phys. Rev. **118**, 1417 (1960).

A MATRIX METHOD FOR INVESTIGATION OF TWO-DIMENSIONAL HEAT CONDUCTION PROBLEMS

BY

VICTOR LOVASS-NAGY

Budapest, Hungary

(Communicated by Prof. W. T. KOITER at the meeting of April 29, 1961)

§ 1. *Introduction*

The approximation of the partial differential equation of heat conduction by finite difference equations has been discussed by several authors (for example, see Refs. [1] [4]). These authors replace, as an approximation, the domain in which the heat conduction process will be investigated, by a point-lattice and the derivatives with respect to both the time variable and the coordinates of lattice points by finite difference quotients. The differential equation governing the problem is then reduced to a system of linear simultaneous algebraic equations which is solved either by some sort of numerical procedure or by a graphical method.

The present paper describes a simple matrix method for the exact numerical evaluation of temperature distributions in the equidistant lattice points of a two-dimensional solution domain of a rectangle covered by a rectangular network for the cases in which the initial conditions in the lattice points are determined by arbitrary functions of the time and on the boundary the temperature distribution satisfies the so-called first boundary condition. It will be assumed that the thermal conductivity in the x direction is different from that in the y direction (this occurs, for example, in laminated transformer cores). Further will be taken into account a given generation of heat per unit volume. In contrast to former authors, in the present paper the derivatives with respect to the time are not replaced by finite difference quotients, i.e. the time is regarded as a continuous independent variable and the temperature is regarded as a continuous function of the time. On the other hand, derivatives with respect to the coordinates x and y, respectively, will be approximated by difference quotients. This method of approximation results in a system of first-order linear differential equations, which can be incorporated in a single matrix differential equation. This matrix differential equation (which includes the boundary conditions too), can be solved by pure quadrature and the solution can be expressed in terms of scalar functions by making use of the canonical representation of the coefficient matrix.

The problem of heat conduction in an irregular domain can also be worked out by the above method. This and the related problems for the so-called second and third boundary conditions will be presented in subsequent papers.

§ 2. *List of Symbols*

a, b, c . . . . . scalars

$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \cdot & \cdot & \dots & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{bmatrix}$ . . . . square matrix composed of the scalar elements a_{ij};

$\mathbf{A}$ $\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1m} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \dots & \mathbf{A}_{2m} \\ \cdot & \cdot & \dots & \cdot \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \dots & \mathbf{A}_{mm} \end{bmatrix}$ . . . square hypermatrix composed of the square submatrices $\mathbf{A}_{ij}$;

$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a \end{bmatrix}$ . . . . . column matrix composed of the scalar elements a_i;

$\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$ . . . . . column hypermatrix composed of the column submatrices $\mathbf{a}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}$;

$\mathbf{A}^*$ . . . . . transpose of $\mathbf{A}$;

$\mathbf{a}^* = [a_1, a_2, \dots, a_m]$ . . . . . row matrix composed of the scalar elements a_i (transpose of the column matrix $\mathbf{a}$);

$[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \cdot & \cdot & \dots & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{bmatrix}$ . . . row hypermatrix composed of the column submatrices $\mathbf{a}_i$;

$\mathbf{E}^{(n)}$ . . . . . unit matrix of the order n;

$\langle d_1, d_2, \dots, d_n \rangle = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & d_n \end{bmatrix}$ . . . diagonal matrix of the order n;

$\langle \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n \rangle = \begin{bmatrix} \mathbf{D}_1 & 0 & \dots & 0 \\ 0 & \mathbf{D}_2 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \mathbf{D}_n \end{bmatrix}$ . . . diagonal hypermatrix;

$$\mathbf{C}^{(n)} = \left[\begin{array}{rrrrrrr} 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 2 & -1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & \dots & 0 & -1 & 2 \end{array}\right] \quad \dots \text{ uniform continuant matrix of the order } n;$$

$$\lambda_k^{(n)} = 4 \sin^2 \frac{k\pi}{2(n+1)}; \quad (k = 1, 2, \dots, n) \quad \dots \text{ eigenvalues of the uniform continuant matrix } \mathbf{C}^{(n)};$$

$$\mathbf{v}_k^{(n)*} = [v_{1k}^{(n)}, v_{2k}^{(n)}, \dots, v_{nk}^{(n)}] = \sqrt{\frac{2}{n+1}} \left[\sin \frac{k\pi}{n+1}, \sin \frac{2k\pi}{n+1}, \dots, \sin \frac{nk\pi}{n+1}\right]$$

$(k = 1, 2, \dots, n)$ . . . . . eigenvectors of the uniform continuant matrix $\mathbf{C}^{(n)}$;

$$\mathbf{A} \times \cdot \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2m}\mathbf{B} \\ \cdot & \cdot & \dots & \cdot \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mm}\mathbf{B} \end{bmatrix} \quad \dots \text{ direct product of the square matrices } \mathbf{A} \text{ and } \mathbf{B};$$

$$\mathbf{a} \times \cdot \mathbf{b} = \begin{bmatrix} a_1\mathbf{b} \\ a_2\mathbf{b} \\ \vdots \\ a_m\mathbf{b} \end{bmatrix} \quad \dots \text{ direct product of the column matrices } \mathbf{a} \text{ and } \mathbf{b};$$

u = temperature within the solution domain;
t = time;
x, y . . . orthogonal coordinates of an internal point;
ξ, η . . . orthogonal coordinates of a boundary point;
ϱ = density;
c = specific heat;
l_x = thermal conductivity in the x direction;
$_y$ = thermal conductivity in the y direction;
$a_x^2 = \dfrac{l_x}{c \cdot \varrho}; \quad a_y^2 = \dfrac{l_y}{c \cdot \varrho};$
q = generation of heat per unit volume;
ϑ = environment temperature.

§ 3. *Deduction of the Governing Matrix Differential Equation of the Problem*

Consider a two-dimensional domain of a rectangle (Fig. 1). It is known that in cases where the density, the specific heat and the thermal conductivities in the x and y directions, respectively, are constant, the temperature distribution within the given domain obeys the following equation:

$$\frac{\partial u}{\partial t} = \frac{1}{c \cdot \gamma}\left[l_x \frac{\partial^2 u}{\partial x^2} + l_y \frac{\partial^2 u}{\partial y^2}\right] + \frac{1}{c \cdot \gamma} q. \tag{1}$$

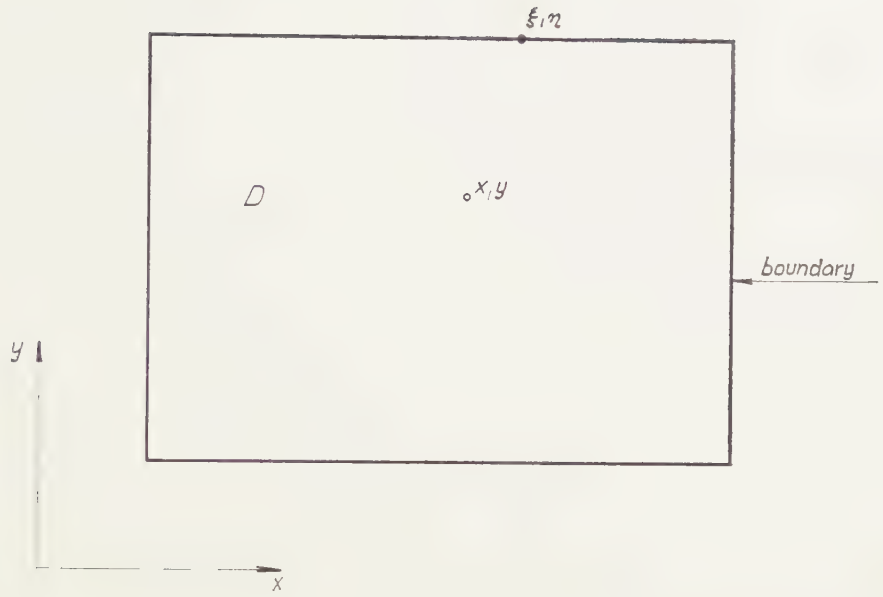

Fig. 1.

(This equation assumes that there is no heat flow in the direction perpendicular to the x, y plane). Further the function $u(x, y; t)$ describing the temperature distribution must satisfy the initial condition

$$u(x, y; 0) = \varphi(x, y) \tag{2}$$

and one of the boundary conditions

$$u(\xi, \eta; t) = \vartheta(\xi, \eta; t) \tag{3a}$$

$$\begin{cases} \dfrac{\partial u}{\partial x}(\xi, \eta; t) = \varrho(\xi, \eta; t) \\ \dfrac{\partial u}{\partial y}(\xi, \eta; t) = \varrho(\xi, \eta; t) \end{cases} \tag{3b}$$

$$\begin{cases} l_x \dfrac{\partial u}{\partial x}(\xi, \eta; t) = \alpha_x(\xi, \eta; t)\,[u(\xi, \eta; t) - \vartheta(\xi, \eta; t)] \\ l_y \dfrac{\partial u}{\partial x}(\xi, \eta; t) = \alpha_y(\xi, \eta; t)\,[u(\xi, \eta; t) - \vartheta(\xi, \eta; t)] \end{cases} \tag{3c}$$

where $\vartheta(\xi, \eta; t)$, $\varrho(\xi, \eta; t)$, $\alpha_x(\xi, \eta; t)$ and $\alpha_y(\xi, \eta; t)$ denote given functions which do not vanish at the same time. The scope of the present paper is restricted to the case where the temperature distribution fulfills the so-called first boundary condition, i.e. Eq. (3a).

Replace the given domain D by a two-dimensional rectangular lattice composed of equidistant lattice-points (as shown in Fig. 2) and the derivatives with respect to the coordinates x and y, respectively, by

difference quotients, the temperature function $u_{ik}(t)$ belonging to the lattice-point with the indices i, k satisfy

$$\left.\begin{aligned}&\frac{d}{dt}\,u_{ik}+\frac{a_x^2}{h^2}\left[-u_{i-1,k}+2u_{ik}-u_{i+1,k}\right]+\\&\quad+\frac{a_y^2}{h^2}\left[-u_{i,k-1}+2u_{ik}-u_{i,k+1}\right]=\frac{1}{c\cdot\gamma}\,q_{ik}\end{aligned}\right\}\tag{4}$$

where h is the distance between two neighbouring lattice-points.

Let us approximate Eq. (1) by the equations (4). Then the initial condition (2) reduces to

$$u_{ik}(0)=\varphi_{ik}\tag{5}$$

where φ_{ik} denotes discrete values of the function $\varphi(x, y)$. The temperature at any point of the boundary of the rectangel satisfies the following equation

$$u_{\mu,\nu}(t)=\vartheta_{\mu,\nu}(t)\tag{6}$$

where

$$\nu=1, 2, \ldots, n \text{ if } \mu=0 \text{ or } \mu=m+1$$

and

$$\mu=1, 2, \ldots, m \text{ if } \nu=0 \text{ or } \nu=n+1.$$

Applying equations (4) and (6) we obtain the following system of first-order linear differential equations:

$$\left\{\begin{aligned}&\frac{d}{dt}\,u_{11}+\frac{a_x^2}{h^2}\left[2u_{11}-u_{21}\right]+\frac{a_y^2}{h^2}\left[2u_{11}-u_{12}\right]-\\&\qquad-\frac{a_x^2}{h^2}\,\vartheta_{01}(t)-\frac{a_y^2}{h^2}\,\vartheta_{10}(t)=\frac{1}{c\cdot\gamma}\,q_{11}\\&\frac{d}{dt}\,u_{21}+\frac{a_x^2}{h^2}\left[-u_{11}+2u_{21}-u_{31}\right]+\frac{a_y^2}{h^2}\left[2u_{21}-u_{22}\right]=\\&\qquad=\frac{a_y^2}{h^2}\,\vartheta_{20}(t)+\frac{1}{c\cdot\gamma}\,q_{21}\\&\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\&\text{and so on.}\end{aligned}\right.\tag{7.1}$$

The above system of equations can be incorporated in the single matrix equation

$$\frac{d}{dt}\,\mathbf{u}+\mathbf{A}\mathbf{u}=\mathbf{g}\tag{7.2}$$

where

$$\mathbf{A}=\frac{a_x^2}{h^2}\,\mathbf{E}^{(n)}\times\cdot\,\mathbf{C}^{(m)}+\frac{a_y^2}{h^2}\,\mathbf{C}^{(n)}\times\cdot\,\mathbf{E}^{(m)}$$

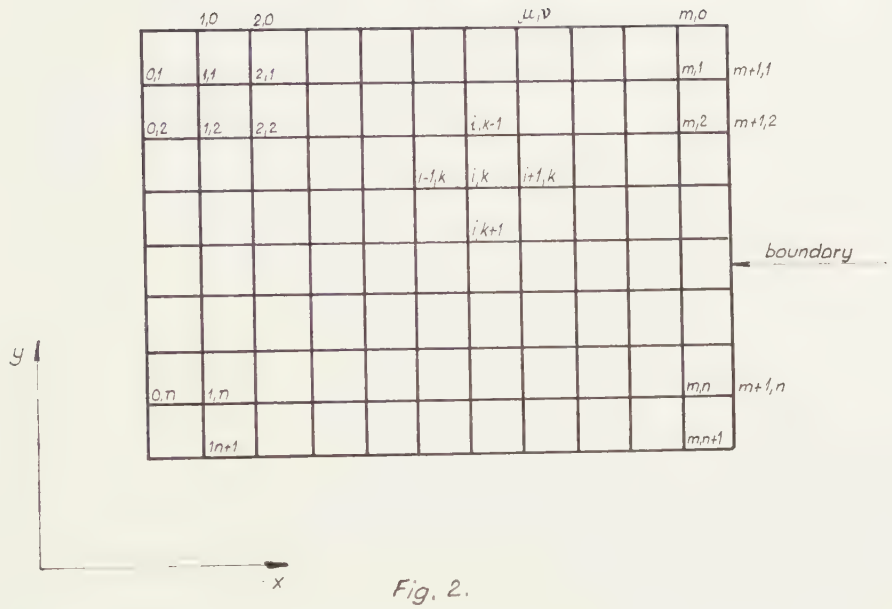

Fig. 2.

and

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{bmatrix}; \quad \mathbf{u}_\nu = \begin{bmatrix} u_{1\nu} \\ u_{2\nu} \\ \vdots \\ u_{m\nu} \end{bmatrix};$$

$$\mathbf{g} = \mathbf{g}' + \mathbf{g}'' + \mathbf{q}$$

$$\mathbf{g}' = \frac{a_x^2}{h^2} \begin{bmatrix} \mathbf{g}_1' \\ \mathbf{g}_2' \\ \vdots \\ \mathbf{g}_n' \end{bmatrix}; \quad \mathbf{g}_\nu' = \begin{bmatrix} \vartheta_{0\nu}(t) \\ 0 \\ \vdots \\ 0 \\ \vartheta_{m+1,\nu}(t) \end{bmatrix}$$

$$\mathbf{g}'' = \frac{a_y^2}{h^2} \begin{bmatrix} \mathbf{g}_1'' \\ 0 \\ \vdots \\ 0 \\ \mathbf{g}_n'' \end{bmatrix}; \quad \mathbf{g}_1'' = \begin{bmatrix} \vartheta_{1,0}(t) \\ \vartheta_{2,0}(t) \\ \vdots \\ \vartheta_{m-1,0}(t) \\ \vartheta_{m,0}(t) \end{bmatrix}; \quad \mathbf{g}_{n+1}'' = \begin{bmatrix} \vartheta_{1,n+1}(t) \\ \vartheta_{2,n+1}(t) \\ \vdots \\ \vartheta_{m-1,n+1}(t) \\ \vartheta_{m,n+1}(t) \end{bmatrix}$$

$$\mathbf{q} = \frac{1}{c \cdot \gamma} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \vdots \\ \mathbf{q}_n \end{bmatrix}; \quad \mathbf{q}_\nu = \begin{bmatrix} q_{1,\nu}(t) \\ q_{2,\nu}(t) \\ \vdots \\ q_{m,\nu}(t) \end{bmatrix}$$

§ 4. *Solution of the matrix differential equation*

Let us introduce the transforming matrix

$$\mathbf{V} = \mathbf{S}^{(n)} \times \cdot \mathbf{S}^{(m)}$$

where

$$\mathbf{S}^{(n)} = [\mathbf{v}_1{}^{(n)}, \mathbf{v}_2{}^{(n)}, \ldots, \mathbf{v}_n{}^{(n)}]$$

$$\mathbf{S}^{(m)} = [\mathbf{v}_1{}^{(m)}, \mathbf{v}_2{}^{(m)}, \ldots, \mathbf{v}_m{}^{(m)}].$$

It is seen that

$$\mathbf{V}^* \quad \mathbf{V}$$

and

$$\mathbf{V} \cdot \mathbf{V} = \mathbf{E}^{(m \times n)}$$

from which it follows that

$$\mathbf{V} \cdot \mathbf{A} \cdot \mathbf{V} = \frac{a_x{}^2}{h^2} (\mathbf{S}^{(n)} \cdot \mathbf{E}^{(n)} \cdot \mathbf{S}^{(n)}) \times \cdot (\mathbf{S}^{(m)} \cdot \mathbf{C}^{(m)} \cdot \mathbf{S}^{(m)})$$

$$+ \frac{a_y{}^2}{h^2} (\mathbf{S}^{(n)} \cdot \mathbf{C}^{(n)} \cdot \mathbf{S}^{(n)}) \times \cdot (\mathbf{S}^{(m)} \cdot \mathbf{E}^{(m)} \cdot \mathbf{S}^{(m)}) =$$

$$= \frac{a_x{}^2}{h^2} \mathbf{E}^{(n)} \times \cdot \mathbf{\Lambda}^{(m)} + \frac{a_y{}^2}{h^2} \mathbf{\Lambda}^{(n)} \times \cdot \mathbf{E}^{(m)}$$

where

$$\mathbf{\Lambda}^{(m)} = \langle \lambda_1{}^{(m)}, \lambda_2{}^{(m)}, \ldots, \lambda_m{}^{(m)} \rangle$$

$$\mathbf{\Lambda}^{(n)} \quad \lambda_1{}^{(n)}, \lambda_2{}^{(n)}, \ldots, \lambda_n{}^{(n)}.$$

Hence an arbitrary analytic function of the matrix $\mathbf{A}$ can be expressed in the form

$$f(\mathbf{A}) = \mathbf{V} \cdot \mathbf{D} \cdot \mathbf{V} \tag{8}$$

where

$$\mathbf{D} = \langle \mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n \rangle; \quad \mathbf{D}_\nu = \langle f(\delta_{1\nu}), f(\delta_{2\nu}), \ldots, f(\delta_{m\nu}) \rangle$$

and

$$\delta_{\mu\nu} = \frac{a_x{}^2}{h^2} \lambda_\nu^{(m)} + \frac{a_y{}^2}{h^2} \lambda_\nu^{(n)}; \quad \mu = 1, 2, \ldots, m; \quad \nu = 1, 2, \ldots, n.$$

Then the solution of (7.2) becomes

$$\mathbf{u} = e^{-\mathbf{A}t} \mathbf{u}^0 + \int_{\tau=0}^{\tau=t} e^{\mathbf{A}(\tau - t)} \mathbf{g}(\tau) \, d\tau \tag{9}$$

where

$$\mathbf{u}^0 = \begin{bmatrix} \mathbf{u}_1^0 \\ \mathbf{u}_2^0 \\ \vdots \\ \mathbf{u}_n^0 \end{bmatrix}; \quad \mathbf{u}_\nu^0 = \begin{bmatrix} u_{1\nu}^0 \\ u_{2\nu}^0 \\ \vdots \\ u_{m\nu}^0 \end{bmatrix}$$

Using the expression (8), it follows from the formula (9) that

$$\mathbf{u} = \sum_{\nu=1}^{n} \sum_{\mu=1}^{m} \mathbf{v}_{\nu}^{(n)} \times \cdot \mathbf{v}_{\mu}^{(m)} \left\{ \sum_{j=1}^{n} \sum_{i=1}^{m} (v_{\nu j}^{(n)}\, v_{\mu i}^{(m)})\, e^{-\delta_{\mu\nu} t}\, u_{ij}^{0} + \right.$$

$$+ \frac{a_x^2}{h^2} \sum_{j=1}^{n} v_{\nu j}^{(n)} \int_{\tau=0}^{\tau=t} e^{\delta_{\mu\nu}(\tau - t)} \left[v_{\mu 1}^{(m)}\, \vartheta_{0j}(\tau) + v_{\mu m}^{(m)}\, \vartheta_{m+1,j}(\tau) \right] d\tau +$$

$$+ \frac{a_y^2}{h^2} \sum_{i=1}^{m} v_{\mu i}^{(m)} \int_{\tau=0}^{\tau=t} e^{\delta_{\mu\nu}(\tau - t)} \left[v_{\nu 1}^{(n)}\, \vartheta_{i0}(\tau) + v_{\nu n}\, \vartheta_{i,n+1}(\tau) \right] d\tau +$$

$$\left. + \frac{1}{c \cdot \gamma} \sum_{j=1}^{n} \sum_{i=1}^{m} (v_{\nu j}^{(n)}\, v_{\mu i}^{(m)}) \int_{}^{\tau=t} e^{\delta_{\mu\nu}(\tau - t)}\, q_{ij}(\tau)\, d\tau \right\}.$$

§ 5. *Appendix*

The method demonstrated in the paragraphs above can be developed for calculation of temperature distributions in irregular domains. The complete working-out of this generalization needs further mathematical investigations, the basic idea of the procedure, however, can be illustrated by means of a simple example.

Let us consider, for example, a solution domain composed of two rectangles (Fig. 3). In this case the governing differential equations can be written in the form

$$\frac{d}{dt}\, \mathbf{u}_{\mathrm{I}} + \mathbf{A}_{\mathrm{I}} \cdot \mathbf{u}_{\mathrm{I}} = \mathbf{g}_{\mathrm{I}}' + \mathbf{g}_{\mathrm{I}}'' + \mathbf{q}_{\mathrm{I}} + \mathbf{K} \cdot \mathbf{u}_{\mathrm{II}} \tag{10a}$$

$$\frac{d}{dt}\, \mathbf{u}_{\mathrm{II}} + \mathbf{A}_{\mathrm{II}} \cdot \mathbf{u}_{\mathrm{II}} = \mathbf{g}_{\mathrm{II}}' + \mathbf{g}_{\mathrm{II}}'' + \mathbf{q}_{\mathrm{II}} + \mathbf{K}^{*} \cdot \mathbf{u}_{\mathrm{I}}. \tag{10b}$$

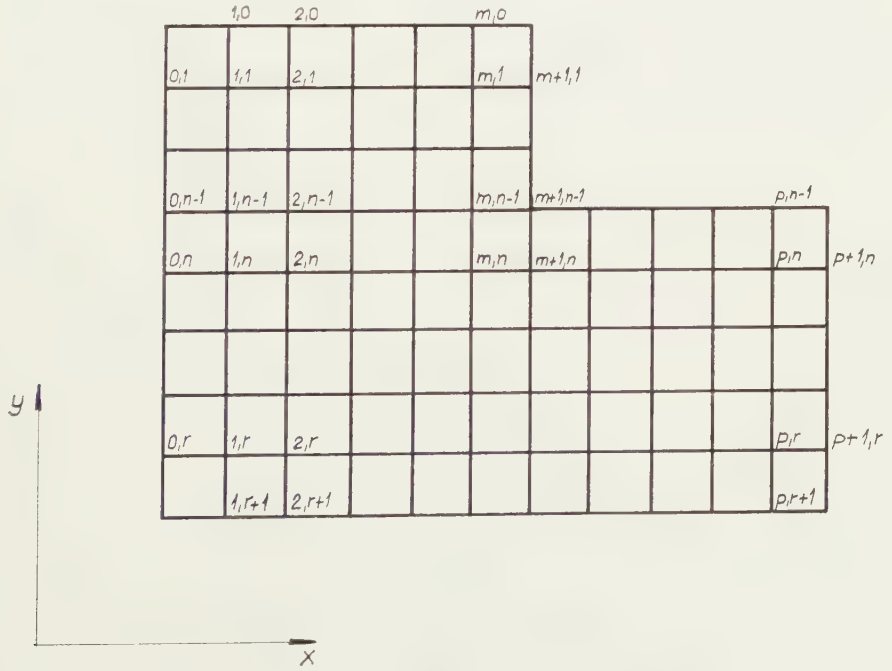

Fig. 3.

Here denotes

$$\mathbf{u}_{\mathrm{I}} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{n-1} \end{bmatrix}; \quad \mathbf{u}_\nu = \begin{bmatrix} u_{1\nu} \\ \vdots \\ u_{m\nu} \end{bmatrix}; \quad \nu = 1, 2, \ldots, n-1$$

$$\mathbf{u}_{\mathrm{II}} = \begin{bmatrix} \mathbf{u}_n \\ \vdots \\ \mathbf{u}_r \end{bmatrix}; \quad \mathbf{u}_\nu = \begin{bmatrix} u_{1\nu} \\ \vdots \\ u_{m\nu} \end{bmatrix}; \quad \nu = n, n+1, \ldots, r$$

$$\mathbf{A}_{\mathrm{I}} = \frac{a_x^2}{h^2}\, \mathbf{E}^{(n-1)} \times \mathbf{C}^{(m)} + \frac{a_y^2}{h^2}\, \mathbf{C}^{(n-1)} \times \mathbf{E}^{(m)}$$

$$\mathbf{A}_{\mathrm{II}} = \frac{a_x^2}{h^2}\, \mathbf{E}^{(r-n)} \times \mathbf{C}^{(p)} + \frac{a_y^2}{h^2}\, \mathbf{C}^{(r-n)} \times \mathbf{E}^{(p)}$$

$$\mathbf{K} \quad \begin{array}{|c|c|} \hline 0 & 0 \\ \hline \mathbf{E}^{(m)} & 0 \\ \hline \end{array} \quad \left.\begin{array}{l} \\ \\ \end{array}\right\} m \times (n-1) \text{ rows}$$

$p = (r - n)$ columns

$$\mathbf{g}_{\mathrm{I}}' = \frac{a_x^2}{h^2} \begin{bmatrix} \mathbf{g}_1' \\ \vdots \\ \mathbf{g}_{n-1}' \end{bmatrix}; \quad \mathbf{g}_\nu' = \begin{bmatrix} \vartheta_{1\nu}(t) \\ \vdots \\ \vartheta_{m+1,\nu}(t) \end{bmatrix}; \quad \nu = 1, 2, \ldots, n-1$$

$$\mathbf{g}_{\mathrm{II}}' = \frac{a_x^2}{h^2} \begin{bmatrix} \mathbf{g}_n' \\ \vdots \\ \mathbf{g}_r' \end{bmatrix}; \quad \mathbf{g}_\nu' = \begin{bmatrix} \vartheta_{1\nu}(t) \\ \vdots \\ \vartheta_{l+1,\nu}(t) \end{bmatrix}; \quad \nu = n, n+1, \ldots, r$$

$$\mathbf{g}_{\mathrm{I}}'' = \frac{a_y^2}{h^2} \left[\begin{array}{c} \mathbf{g}_0'' \\ 0 \\ 0 \\ \vdots \\ 0 \end{array}\right\} n-1 \text{ elements} \qquad ; \quad \mathbf{g}_0'' = \begin{bmatrix} \vartheta_{10}(t) \\ \vartheta_{20}(t) \\ \vdots \\ \vartheta_{m0}(t) \end{bmatrix}$$

$$\mathbf{g}_{\mathrm{II}}'' = \frac{a_y^2}{h^2} \left[\begin{array}{c} \mathbf{g}_{n-1}'' \\ 0 \\ 0 \\ \vdots \\ 0 \\ \mathbf{g}_{r+1}'' \end{array}\right\} r-1 \text{ elements}$$

$$\mathbf{g}''_{\nu-1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vartheta_{m+1,\,n-1}(t) \\ \vdots \\ \vartheta_{p,\,n-1}(t) \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}0\\\vdots\\0\end{matrix}}\right\} m \text{ elements are zero} \\ \left.\vphantom{\begin{matrix}a\\\vdots\\b\end{matrix}}\right\} p-m \text{ elements} \end{matrix}; \quad \mathbf{g}''_{r+1} = \begin{bmatrix} \vartheta_{1,r+1}(t) \\ \vartheta_{2,r+1}(t) \\ \vdots \\ \vartheta_{p,r+1}(t) \end{bmatrix}$$

$$\mathbf{q}_{\mathrm{I}} = \frac{1}{c\cdot\gamma}\begin{bmatrix} \mathbf{q}_1 \\ \vdots \\ \mathbf{q}_{n-1} \end{bmatrix}; \quad \mathbf{q}_\nu = \begin{bmatrix} q_1 \\ \vdots \\ q_m \end{bmatrix}; \quad \nu = 1, 2, \ldots, n-1$$

$$\mathbf{q}_{\mathrm{II}} = \frac{1}{c\cdot\gamma}\begin{bmatrix} \mathbf{q}_n \\ \vdots \\ \mathbf{q}_r \end{bmatrix}; \quad \mathbf{q}_\nu = \begin{bmatrix} q_1 \\ \vdots \\ q_p \end{bmatrix}; \quad \nu = n, n+1, \ldots, r.$$

Let us introduce the notations

$$\mathbf{u} \quad \begin{bmatrix} \mathbf{u}_{\mathrm{I}} \\ \mathbf{u}_{\mathrm{II}} \end{bmatrix}; \ \mathbf{g} - \begin{bmatrix} \mathbf{g}_{\mathrm{I}}' \\ \mathbf{g}_{\mathrm{II}}' \end{bmatrix} \quad \begin{bmatrix} \mathbf{g}_{\mathrm{I}}'' \\ \mathbf{g}_{\mathrm{II}}'' \end{bmatrix}; \ \mathbf{A} - \begin{bmatrix} \mathbf{A}_{\mathrm{I}} & 0 \\ 0 & \mathbf{A}_{\mathrm{II}} \end{bmatrix}; \ \mathbf{B} - \begin{bmatrix} 0 & \mathbf{K} \\ \mathbf{K}^* & 0 \end{bmatrix}.$$

Then equations (10a) and (10b) reduce to

$$\frac{d}{dt}\,\mathbf{u} = -\,\mathbf{A}\mathbf{u} - \mathbf{B}\mathbf{u} + \mathbf{g}(t). \tag{11}$$

The initial condition for $t=0$ is $\mathbf{u}(0)=\mathbf{u}_0$.

In a former article [5] by the author of the present paper there was given a method for finding an approximate solution of Eq. (11), satisfying the initial condition $\mathbf{u}(0)=\mathbf{u}_0$. Since the spectral decompositions of the matrices $\mathbf{A}_{\mathrm{I}}$ and $\mathbf{A}_{\mathrm{II}}$ are explicitely known, an arbitrary analytic function of $\mathbf{A}$ can be written in the form

$$f(\mathbf{A}) = \langle \mathbf{V}_{\mathrm{I}}, \mathbf{V}_{\mathrm{II}} \rangle \langle \mathbf{D}_{\mathrm{I}}, \mathbf{D}_{\mathrm{II}} \rangle \langle \mathbf{V}_{\mathrm{I}}, \mathbf{V}_{\mathrm{II}} \rangle$$

where

$\mathbf{V}_{\mathrm{I}} = \mathbf{S}^{(n-1)} \times \cdot\, \mathbf{S}^{(m)}$

$\mathbf{V}_{\mathrm{II}} = \mathbf{S}^{(r-n)} \times \cdot\, \mathbf{S}^{(p)}$

$\mathbf{D}_{\mathrm{I}} = \langle \mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{n-1} \rangle$; $\mathbf{D}_\nu = \langle f(\delta_{1\nu}), f(\delta_{2\nu}), \ldots, f(\delta_{m\nu}) \rangle \quad \nu = 1, 2, \ldots, n-1$;

$\mathbf{D}_{\mathrm{II}} = \langle \mathbf{D}_n, \mathbf{D}_{n+1}, \ldots, \mathbf{D}_r \rangle$; $\mathbf{D}_\nu = \langle f(\delta_{1\nu}), f(\delta_{2\nu}), \ldots, f(\delta_{p\nu}) \rangle \quad \nu = n, n+1, \ldots, r.$

It was shown in [5], that the sequence of the solutions of the equations

$$\frac{d}{dt}\,\mathbf{u}_{h1} = -\,\mathbf{A}\mathbf{u}_{h1} \qquad ; \ \mathbf{u}_{h1}(0) = \mathbf{u}_0$$

$$\frac{d}{dt}\,\mathbf{u}_{h2} = -\,\mathbf{A}\mathbf{u}_{h2} - \mathbf{B}\mathbf{u}_{h1}; \ \mathbf{u}_{h2}(0) = \mathbf{u}_0$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\frac{d}{dt}\,\mathbf{u}_{h\nu} = -\,\mathbf{A}\mathbf{u}_{h\nu} - \mathbf{B}\mathbf{u}_{h,\nu-1}; \ \mathbf{u}_{h\nu}(0) = \mathbf{u}_0$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

has as its limiting value an exact solution of the homogeneous equation associated with Eq. (11), i.e.

$$\lim_{\nu \to \infty} \mathbf{u}_{h\nu} = e^{-(\mathbf{A}+\mathbf{B})t}\, \mathbf{u}_0.$$

It is clear that $\mathbf{u}_{h\nu}$ may be expressed in the form

$$u_{h\nu} = e^{-\mathbf{A}t}\, \mathbf{u}_0 - \int_{\tau=0}^{\tau=t} e^{-\mathbf{A}(t-\tau)}\, \mathbf{B}\mathbf{u}_{h,\nu-1}(\tau)\, d\tau =$$

$$= \left\{ e^{-\mathbf{A}t} - \int_{\tau=0}^{\tau=t} e^{-\mathbf{A}(t-\tau)}\, \mathbf{B}\mathbf{M}_{\nu-1}(\tau)\, d\tau \right\} \mathbf{u}_0 = \mathbf{M}_\nu(t)\, \mathbf{u}_0$$

where $\mathbf{M}_0(t)$ is a zero matrix. Hence an approximate solution of the $(\nu-1)$th order of Eq. (11) is given by

$$\mathbf{u}_\nu = \mathbf{M}_\nu(t)\, \mathbf{u}_0 + \int_{\tau=0}^{\tau=t} \mathbf{M}_\nu(t-\tau)\, \mathbf{g}(\tau)\, d\tau. \tag{12}$$

Abstract

This paper presents what is believed to be a new approach to the problem of determining temperature distributions in two-dimensional domains. The method of analysis adapted is as follows: The continuous solution domain D will be replaced by a lattice of discrete points, and, instead of obtaining a continuous function of the co-ordinates x and y for the temperature defined throughout D, we find approximations to the temperature only at these isolated points. Consequently the derivatives with respect to x and y will be approximated by expressions of finite differences. The time t, however, will be regarded as a continuous variable, and the derivatives with respect to t will not be replaced by finite difference quotients. This method of approximation results in a system of first-order differential equations. This system of equations can be incorporated into a single matrix equation, which includes the boundary conditions too. The initial conditions may be arbitrary. A simple matrix method is given for the case in which the solution domain is a rectangle, and on the boundary the temperature distribution satisfies the so-called first boundary condition. Advantage is taken of the fact, that the coefficient matrix of the equation can be expressed by means of the so called uniform continuant matrices, the eigenvalues and eigenvectors of which are explicitely known for arbitrary order. In the appendix some suggestions are given as to the extension of the method for the calculation of temperature distributions in irregular domains.

REFERENCES

1. Schmidt, E., "Über die Anwendung der Differenzenrechnung auf technische Mechanik und technischen Physik. (August Föppl Festschrift) Verlag Anheiz- und Abkühlungsprobleme." Beiträge zur technischen von J. Springer, Berlin, 179–189 (1924).

2. Richter, A.: "Zweidimensionale nichtstationäre Felder der Wärmeleitung." Archiv für Elektrotechnik, **41**, Heft 5, 258–281 (1954).
3. Panow, D. J , Formelsammlung zur numerischen Behandlung partieller Differentialgleichungen nach dem Differenzenverfahren. Akademie-Verlag, Berlin (1955).
4. Crandall, S. H., Engineering Analysis. McGraw-Hill Book Co., New York/Toronto/London (1956).
5. Bajcsay, P. and V. Lovass-Nagy, "Ein Iterationsverfahren zur näherungsweisen Lösung von Matrizendifferentialgleichungen." Zeitschrift für angewandte Mathematik und Mechanik, **39**, 8–13 (1959).

PATTERN OF PLAGIOCLASE TWINNING AS A SIGNIFICANT ROCK PROPERTY

BY

A. C. TOBI

(Communicated by Prof. H. A. Brouwer at the meeting of June 26, 1961)

Introduction

After the development of the universal stage an extensive study was made of the modes of plagioclase twinning in various rock types. Although some of the results seemed promising, this study did not succeed in becoming an important tool in petrology. This was probably due to the rather complicated determination procedure and the large number of twin laws making difficult the distinction of the rules underlying their occurrence.

To overcome this difficulty, Gorai (1951) divided the various twin laws of plagioclase into two groups claimed to be distinguishable without the aid of the universal stage. One group, consisting of the usually lamellar albite, acline and pericline laws, were designated as A-twins, all other laws as C-twins. Lamellar twins with two different extinction positions for X′ are A-twins, lamellar twins with three or four different extinction positions, being of composite nature, are C-twins. Simple twins consisting of two individuals only are often difficult to identify. The relative amounts of A- and C-twins and of untwinned crystals (U) were plotted by Gorai in UAC triangles. It appeared from his study that only magmatic rocks showed any appreciable amount of C-twins. This amount proved to be also dependent upon the An-content, basic rocks having more C-twins than acid rocks.

Although Gorai's method has proved very valuable in petrography, in the opinion of the author the composition plane is at least as important as the type of twinning. To illustrate this opinion, the characteristic patterns of twinning found in some major rock groups are here briefly summarized.

Magmatic rocks

Many twins of all kinds, the patterns showing variations from rock to rock. C-twins are present, and more numerous in basic than in acid rocks. *As a general rule, twins with* (010) *as composition plane are far more frequent than other twins.*

Thermometamorphic rocks

The relative amount of twinned crystals is smaller than in magmatic rocks, while C-twins are usually absent (Gorai). *On the other hand, the pattern of twinning agrees with that of magmatic rocks in the relative abundance of* (010) *twins.*

Metamorphic rocks in greenschist facies

Usually few twinned crystals, but sometimes more than in hornfelses. *No lamellar twinning, though simple albite, Carlsbad and albite-Carlsbad twins are all found. In this respect these rocks do not follow the rule of Gorai.* The plagioclase being albite in this facies, the distinction between albite and Carlsbad twins is difficult even with the universal stage. Quite characteristic is the occurrence of simple twins with a narrow central lamella or wedge of other extinction (fig. 1). *Twins with composition planes*

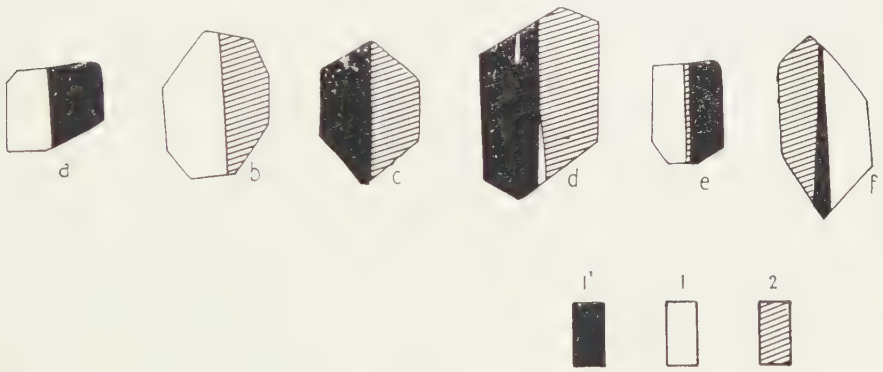

Fig. 1. Characteristic twinning in albite porphyroblasts of the greenschist facies. 1—1': albite law; 1—2: Carlsbad law; 1'—2: albite-Carlsbad law. Usually a, c and d are the most frequent types.

other than (010) *are rigorously absent* [1]). These rules are especially applicable to albite porphyroblasts. In any case the albite must have *crystallized* during greenschist facies conditions, and not during an earlier stage in the history of the rock.

Metamorphic rocks in almandine-amphibolite facies

A rather large amount of twinned crystals, showing exclusively lamellar A-twins, *the acline and pericline laws being more frequent than the albite law.* This fact was also brought forward by Turner (1951). The treatment of this group as a whole implies that chemical and mineralogical composition are of secondary importance for the pattern of twinning: amphi-

[1]) With the exception of the typical quadruplets found in limestones and consisting of the albite (010) and Carlsbad–X (100) laws. These, however, are chiefly authigenic (Fuchtbauer 1956).

bolites and micaschists show the same characteristics in this respect. It is tentatively suggested that preponderance of (010) twins in medium-grade metamorphic rocks points to thermal influences. Thus, the schists and gneisses studied by Gorai, for example, with their high amount of (010) twins, would belong to the non-almandinous (Buchan) rather than to the almandinous (Dalradian) type of metamorphism.

Quantitative analysis: the (010) *twin ratio*

We have shown the relative content of (010) twins to be as important as that of C-twins. A quantitative analysis, leading to a determination of the (010) twin ratio, is only possible with the universal stage. In Table I the ratios of different rock types are given: it is based upon twin analyses of Gorai, Chapman (1936) and the author. The maximum value is reached in the greenschist facies, the minimum value in the almandine-amphibolite facies. The magmatic and associated rocks show comparatively little variation. Here, the higher values (about 0.9) are found in the groundmass of volcanics, the lower ones (about 0.7) in basic thermometamorphic rocks. As stated above, the schists and gneisses studied by Gorai are tentatively assumed to belong to the Buchan type of metamorphism. The extra low value for very basic plagioclase in this group (0.33) constitutes a problem.

On the whole, the influence of the An-content on the (010) twin ratio appears to be limited. The general tendency is therefore best shown by the total values.

Qualitative analysis

A quantitative study has its difficulties, because it implies use of the universal stage, and also because twins cut parallel to the composition plane are not visible as such. It is therefore important to note that the characteristic patterns of twinning described above differ in a qualitative rather than in a quantitative way. They became evident not only from analyses as listed in Table I, but also from a qualitative study of a large number of thin sections from various regions of the world. We should ask the following questions:

1) Are C-twins present?
2) Are twins with other composition planes than (010) present?
3) If so, do they dominate over the (010) twins or not?

The characteristic qualitative differences here discussed, are schematically given in fig. 2.

Recognition of plagioclase twins

For clarity, the main laws of plagioclase twinning are brought here together into the four distinguished twin groups. Complex twins are

TABLE I

The (010) twin ratios of various rock types. The number of twins counted is indicated between brackets.

$\frac{010\ tw}{\Sigma\ tw}$	An per cent	0–24	25–49	50–74	75–100	Total	
greenschist facies (albite porphyroblasts)		1.00 (∞)	—	—	—	1.00 (∞)	3.
volcanic and associated rocks	groundmass	1.00 (12)	0.92 (47)	0.94 (139)	0.89 (82)	0.92 (280)	1.
	phenocrysts	0.80 (10)	0.77 (179)	0.80 (176)	0.83 (150)	0.80 (515)	1.
differentiated sill (diabase)		—	0.89 (136)	0.89 (1273)	—	0.89 (1409)	2.
plutonic rocks		0.91 (53)	0.77 (82)	0.73 (248)		0.76 (383)	1.
hornfelses		—	0.77 (47)	0.76 (70)	0.65 (17)	0.75 (134)	1.
schists and gneisses (Buchan type?)		—	0.75 (84)	0.75 (24)	0.33 (43)	0.63 (151)	1.
almandine-amphibolite facies	micaschist	—	0.19 (138)	—	—	0.19 (138)	4.
	amphibolite	—	0.36 (45)	—	—	0.36 (45)	5.
	total	—	0.23 (183)	—	—	0.23 (183)	

[1] Japan, Gorai (1951).
[2] USA, Chapman (1936).
[3] Everywhere observed in albite porphyroblasts, Tobi (1959).
[4] Belledonne Massif, France, sample B 673 b, Tobi (1959).
[5] „ „ „ „ B 698, Tobi (1959).

omitted, normal twins, having symmetrical extinctions, are printed in italics.

	A (usually lamellar)	C (usually simple)
Composition plane (010)	*albite*	Carlsbad Ala B
Other composition planes	acline pericline	*Manebach* Ala A *Baveno*

The possibilities of twin recognition without the universal stage are discussed in another paper (Tobi, in press). May a few examples here be given. The Baveno twin is recognized by its oblique composition plane. The distinction between the composition planes (010) and (001) is usually only possible when the An-content is below 70 %. In this range the elongation with regard to (010) is always negative: each twin with positive elongation must have another composition plane. Lamellar twins with negative elongation cut normal to the composition plane are also easily identified: asymmetrical extinctions then point to the acline or pericline twin, symmetrical extinctions to the albite twin.

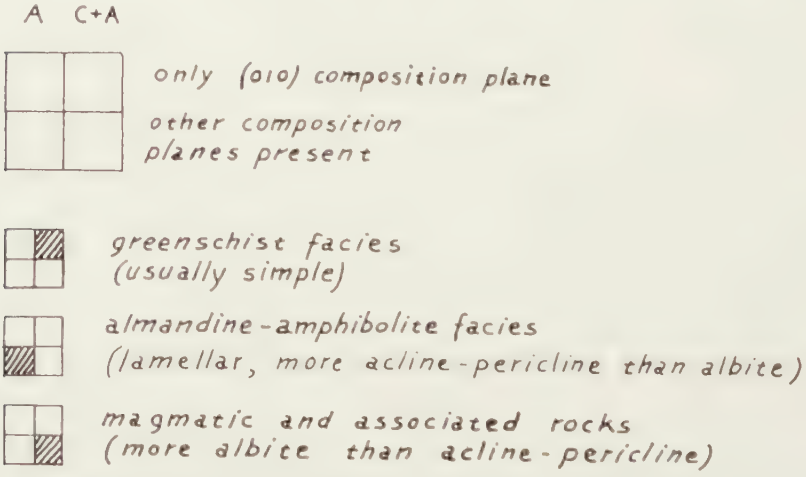

Fig. 2. Diagrams showing qualitative differences in the plagioclase twinning pattern of some major rock groups. The presence or absence of C-twins and of composition planes other than (010) is indicated by shading of a quarter of the squares the significance of which is shown in the large square.

Concluding remarks

Evidently, the patterns of plagioclase twinning show characteristic differences in different rock types. The large number of twin laws and the complicated determination methods call for a simplified procedure.

Apart from the very useful distinction between A- and C-twins proposed by Gorai, the author suggests that the (010) twin ratio should also be taken into consideration. He feels confident that an extensive study of plagioclase twinning will lead to important petrological conclusions.

It should be pointed out that the twinned plagioclase crystals found in metamorphic rocks can be of clastic origin or be inherited from a magmatic or former metamorphic phase. It is possible that new twins are added to a plagioclase crystal during a later phase, but already existing twins are maintained unless complete recrystallization takes place. To trace former phases in polymetamorphic rocks, the pattern of twinning will often be more important than the An-content, which is liable to easy alteration.

For similar reasons, the study of patterns of twinning may be especially helpful in determining the provenance of plagioclase material in arkoses and greywackes.

REFERENCES

Chapman, W. M., A study of feldspar twinning in a differentiated sill. Amer. Miner. **21**, 33–47 (1936).

Fuchtbauer, H., Zur Entstehung und Optik authigener Feldspäte. N. Jahrb. Miner., Mon. h., 9–23 (1956).

Gorai, M., Petrological studies on plagioclase twins. Amer. Miner. **36**, 884–901 (1951).

Tobi, A. C., Petrographical and geological investigations in the Merdaret–Lac Crop region (Belledonne Massif, France). Leidse geol. Meded. **24**, 181–282 (1959).

————, (in press), The recognition of plagioclase twins in sections normal to the composition plane. (Amer. Miner.).

Turner, F. J., Observations on twinning of plagioclase in metamorphic rocks. Amer. Miner. **36**, 581–589 (1951).

CONTENTS

PRINTED IN THE NETHERLANDS
DRUKKERIJ HOLLAND N.V., AMSTERDAM